

M-estimation with probabilistic models of geodetic observations

Z. Wiśniewski

Received: 10 January 2014 / Accepted: 23 May 2014 / Published online: 20 June 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract The paper concerns *M*-estimation with probabilistic models of geodetic observations that is called $M_{\mathcal{P}}$ estimation. The special attention is paid to $M_{\mathcal{P}}$ estimation that includes the asymmetry and the excess kurtosis, which are basic anomalies of empiric distributions of errors of geodetic or astrometric observations (in comparison to the Gaussian errors). It is assumed that the influence function of $M_{\mathcal{P}}$ estimation is equal to the differential equation that defines the system of the Pearson distributions. The central moments μ_k , $k = 2, 3, 4$, are the parameters of that system and thus, they are also the parameters of the chosen influence function. The $M_{\mathcal{P}}$ estimation that includes the Pearson type IV and VII distributions ($M_{PD(I)}$ method) is analyzed in great detail from a theoretical point of view as well as by applying numerical tests. The chosen distributions are leptokurtic with asymmetry which refers to the general characteristic of empirical distributions. Considering *M*-estimation with probabilistic models, the Gram–Charlier series are also applied to approximate the models in question (M_{G-C} method). The paper shows that $M_{\mathcal{P}}$ estimation with the application of probabilistic models belongs to the class of robust estimations; $M_{PD(I)}$ method is especially effective in that case. It is suggested that even in the absence of significant anomalies the method in question should be regarded as robust against gross errors while its robustness is controlled by the pseudo-kurtosis.

Keywords Geodetic adjustment · *M*-estimation · Probabilistic models · Pearson density functions · $M_{\mathcal{P}}$ estimation

1 Introduction and motivation

When adjusting geodetic observations, anomalies (in relation to the conventional assumptions) are more often considered. The expected anomalies concern probability distributions and/or moments of measurement errors. In this context, for example, the observations which are affected by gross errors are considered. Such disturbances are identified in two basic ways. The first way assumes that gross errors are regarded as random variables which have the same variance as other measurement errors but differ from them in expected values. Thus, identification is carried out before the least squares adjustment and by applying various testing approaches (e.g., Baarda 1968; Prószyński 1997; Cen et al. 2003). Another approach is to use a robust method of estimation, including *L*-estimation or *R*-estimation (Huber 1981; Hodges and Lehmann 1963; Duchnowski 2011, 2013); L_1 -norm estimation (Marshall 2002) or *M*-estimation (Huber 1981). While applying robust *M*-estimation, it is usually assumed that gross errors are realizations of random variables which differ from other measurement errors in variances.

Anomalies which concern the expected value and/or the variance are however not the only ones which can be expected in the sets of geodetic observations. The problem can concern the higher order moments and hence asymmetry and excess coefficients (namely, anomalies in relation to the normal distribution). In such a case, usage of *M*-estimation with an adequate probabilistic model of measurement errors gives the opportunity to estimate the parameters of the functional model.

If the vector of the independent observations \mathbf{y} is described by the functional model $\mathbf{y} = \mathbf{F}(\mathbf{X}) + \mathbf{v}$, *M*-estimation can be considered as the solution of the optimization problem $\min_{\mathbf{X}} \phi(\mathbf{y}; \mathbf{X}) = \phi(\mathbf{y}; \hat{\mathbf{X}})$ with the objective function

Z. Wiśniewski (✉)
Institute of Geodesy, University of Warmia and Mazury,
1 Oczapowskiego St., 10-957 Olsztyn, Poland
e-mail: zbyszekw@uwm.edu.pl

$$\phi(\mathbf{y}; \mathbf{X}) = \sum_{i=1}^n \rho(y_i; \mathbf{X}) = \sum_{i=1}^n \rho(v_i) \quad (1)$$

(e.g., Huber 1981; Hampel et al. 1986). Here $v_i = y_i - F_i(\mathbf{X})$ is a random error of the observation. The error distribution P_{θ_i} is indexed with the parameter $\theta_i \in \Theta$ (Θ is a space of the parameter). The class of the distributions $\mathcal{P} = \{P_{\theta_i} : \theta \in \Theta\}$ will be regarded as a probabilistic model of the observation error, and after some suitable rearrangements of the parameters, as a probabilistic model of the observation. The functions $\rho(y_i; \mathbf{X}) = \rho(v_i)$ are arbitrary components of the objective function $\phi(\mathbf{y}; \mathbf{X})$ and usually only in general terms they are referred to the probabilistic models of the observations. However, the probabilistic models are strongly related to the background of M -estimation, namely the maximum likelihood method (ML-method, see, e.g., Serfling 1980; Huber 1981; Winkelmann and Boes 2006).

One should notice that, for example, the least squares method (LS-method) also belongs to the class of M -estimation. In such case $\rho(y; \mathbf{X}) = \rho(v) = pv^2$. Here, $p \propto \mu_2^{-1}$ is the weight of the observation related to the variance $\mu_2 = \sigma^2$ which is the central moment $\mu_k = \mu_k(v) = E\{[v - E(v)]^k\}$ for $k = 2$ (σ is the standard deviation, $E(\circ)$ is the expected value). On the other hand, the subclass of robust M -estimates is of great interest (Huber 1981; Hampel et al. 1986; Serfling 1980). Application of that subclass in geodetic adjustment is analyzed, for example, in Krarup and Kubik 1983; Yang 1994; Huang and Mertikas 1995; Gui and Zhang 1998. Generally, robust M -estimates are determined by using the function $\rho(v) = \hat{p}v^2$, where $\hat{p} \propto \hat{\mu}_2^{-1}$ is the equivalent weight ($\hat{\mu}_2 = \hat{\sigma}^2$ is the equivalent variance; see, e.g., Krarup and Kubik 1983; Yang 1994, 1999; Koch 1996).

Considering M -estimation, which applies the presented components of the objective function, only two moments of v should be determined, namely $E(v)$ and μ_2 (usually $E(v) = 0$). Thus, the classes of distributions indexed with the parameter $\theta \in \{E(v), \mu_2\}$ are the implicit probabilistic models. The class of symmetric and mesokurtic distributions is a very important group here. It consists of distributions for which the asymmetry coefficient $\beta_1 = \mu_3/\mu_2^{3/2}$ is equal to zero and the kurtosis $\beta_2 = \mu_4/\mu_2^2$ is equal to 3 (within the paper we will also use the asymmetry coefficient defined as $\gamma_1 = \text{sgn}(\mu_3)\beta_1^{1/2}$ and the excess $\gamma_2 = \beta_2 - 3$). Note, that $\beta_1 = 0$ and $\beta_2 = 3$ if a random variable is normally distributed, $ND[E(v), \mu_2]$.

Applying such M -estimation we should hope that the values of the higher moments are close to the values for the normal distribution, namely $\mu_3 = \mu_5 \cong 0$, $\mu_4/\mu_2^2 \cong 3$ etc. (similar assumptions might also concern other robust methods, see, for example, Duchnowski 2013). Such hope can be based on the Gaussian theory of measurement errors; however, analyses of observations, among others geodetic mea-

surements, show that this is not always justified. The non-zero asymmetry and excess may also concern the algorithms for the Gaussian white noise simulations. For example, the algorithm that was analyzed in Hu et al. (2001) showed the kurtosis $\beta_2 = 3.10 \div 3.15$ (it can be checked that the simulator of the Gaussian random numbers in MatLab is also not free of such anomalies; which will be used in the empirical part of the paper).

Determination of deviations from the theoretical values of the moments is especially important in the case of precise observations. For example, Dzhan' (1992) showed that every astronomical instrument gives measurement errors with a certain kurtosis, which is mostly close to $\beta_2 = 3.8$. In contemporary astrometric experiments the kurtosis is even bigger, for example, within the project MERIT $\beta_2 = 4.858$ (Dzhan' 2012). Here, the asymmetry coefficient is equal to $\beta_1 = 0.0048$. The high value of the kurtosis means that the random errors of the measurements are concentrated around zero more than in the case of the Gaussian errors. Winter (1978) obtained high positive excess when he analyzed the distribution of measurements of a precise distance meter. Similar property can also be observed for other measurement errors, e.g., Wassef (1959) obtained $\beta_2 = 3.8$ for precise leveling and Kukuća (1967) for hydrostatic leveling ($\beta_1 = 0.012$, $\beta_2 = 4.325$).

Description of measurement errors only by the second moments (or their estimates, for example, by RMS) is also not enough in relation to modern techniques of measurements, like the satellite laser ranging (SLR) or the global positioning system (GPS). Hu et al. (2001) analyzed normalized residuals of SLR observations for Lageos-2 from 13 stations. The kurtosis varied within the interval $\beta_2 = 2.69 \div 9.46$ which depended on the range of the additional parameters (e.g., difference of orbits, the Earth orientation parameters). It is worth noting that only in one of the six cases, the kurtosis obtained value < 3 . The similar values, namely $\beta_2 = 2.98 \div 9.26$, were obtained for the regional GPS network (GPS data of 10 sites, Hu et al. 2001). The authors of both analyses did not compute the asymmetry coefficients, however, the values obtained for the third moment $\mu_3 = -1.51 \div 0.01$ (for SLR data) and $\mu_3 = -0.06 \div 0.12$ (for GPS data) showed the slight asymmetry especially in the first case. Luo et al. (2011) presented the analysis of the influences of different factors on the probability distribution of GNSS observables (phase measurements from the SAPOS®). The analysis concerned the representative studentized double difference residuals (SDDR) and the results are consistent with the presented in the previous paragraph. For 268 SDDR time series (5 baselines) and for the significance level $\alpha = 0.01$, the authors obtained $\gamma_1 = -0.11 \div 0.11$ and $\beta_2 = 2.79 \div 3.29$ (with the predominance of $\beta_2 > 3$).

The review of the kurtosis and asymmetry of the empirical distributions of the observations shows that the distrib-

utions are generally leptokurtic ($\beta_2 > 3$) with more or less significant asymmetry ($\gamma_1 \neq 0$). Romanowski and Green (1983) showed that the leptokurtosis is a dominant property of real sets of observations, especially of high precision ones. Platykurtic distributions ($\beta_2 < 3$) are also possible, however, they are very rare.

Approximation of empirical distributions of measurement errors, which kurtosis and asymmetry are known, can be conveniently done by applying theoretical distributions which are “controlled” by the coefficients β_1 and β_2 . To choose such distribution one can use probability estimation methods. Among others, one can apply the methods based on the Pearson distributions (Pearson 1920), the saddlepoint approximation (Daniels 1954), maximum entropy principle (MEP) (Jaynes 1957) and the Johnson system (Johnson 1949). The extensive comparative analysis which is presented in (Xi et al. 2012) shows that none of those methods has an overall advantage over others. However, the Pearson system presents better accuracy for moderate asymmetry and for a wide range of kurtosis. Note that for large asymmetry the method which gains the advantage is MEP (probabilistic density function approximated by maximizing the entropy subject to the known moments). We showed previously that one should not expect such large anomalies for geodetic observation sets. Thus, but also on the basis of the previous papers of the author (Wiśniewski 1985, 1987, 1991), we will assume that the system of the Pearson distributions (PD-system) which contains types I ÷ XII (Pearson 1920; Elderton 1953; Friori and Zenga 2009; Xi et al. 2012) is the set of theoretical distributions which are steered by the four first moments.

The density functions of the PD-system have different shapes: skewed or symmetric, limited or unlimited range, bell-, U-, J-shaped (including the Gaussian bell curve as a boundary case). It is obvious that only a few types of PD-system can be applied in the theory of measurement errors. Symmetric leptokurtic distributions are described by the Pearson distributions of the type VII (PVII). Such distributions were applied, for example, to analyze errors of precise levelling (Wassef 1959) and to approximate observation errors in astrometry (Dzhun’ 1969, 1991, 1992). PVII were also the basis for the analysis of how kurtosis $\beta_2 > 3$ affects LS-estimates (Wiśniewski 1985). The Pearson distributions of the type IV (PIV) are asymmetric leptokurtic distributions which under $\beta_1 \rightarrow 0$ tends to PVII. Analyzing empirical distributions of geodetic measurement errors one could also consider type I ($\beta_1 > 0$, $\beta_2 < 3$) and its boundary case, namely type II ($\beta_1 = 0$, $\beta_2 < 3$). However, there are some problems which result from the limited range of the random variable as well as the resulting U-shaped weight function (Wiśniewski 1987, 1989, 1991). Fortunately empirical distributions of observations are rarely platykurtic (Romanowski and Green 1983). Thus, such distributions will not be discussed within the paper in detail.

For small asymmetry and moderate excess one can consider replacing a theoretical distribution by its expansion derived from the normal distribution. This can be done by applying the expansion with respect to orthogonal polynomials which is called the Gram–Charlier of the type A (see, e.g., Elderton 1953; Kolassa 2006; Berberan-Santos 2007).

It is important to know how non-normal distribution influences the estimates obtained. The problem is a general one and from such point of view it is considered, for example, in Mooijaart (1985); Klein and Moosbrugger (2000); Mukhopadhyay (2005). It is also an issue discussed in the applied statistics, for example, in marketing Andreassen et al. (2006) and in econometrics Nagahara (2011). The influence of deviation of an empirical distribution from the normal distribution on LS-estimates in the case of geodetic networks was discussed in Gleinsvik (1971, 1972) and Wiśniewski (1985). It was shown that the asymmetry of the distribution, which changed the expected value of the measurement errors, also influenced the bias of the LS-estimates of the parameters of the observation models. Such influences are reinforced or weakened by the kurtosis.

To consider the asymmetry or excess of observations during an adjustment procedure it is necessary to assume a certain probabilistic model of observations, which can be acceptable. Such model results in a particular optimization problem of ML-method, or equivalent particular influence and weight functions of M -estimation. Within this paper we assume that leptokurtic distributions with asymmetry are described by the Pearson distribution of the types IV or VII; or are approximated by the Gram–Charlier series (G–C series). The optimization problem of ML-method with the application of the Pearson distributions was considered in my previous papers (Wiśniewski 1987, 1989, 1991). Dumalski and Wiśniewski (1994) proposed a method of adjustment based on G–C series. The general form of the weight function, which is related to the systems of the Pearson distributions, was presented in Dzhun’ (2011). We will propose a new variant of M -estimation which depends on the variance but also on the kurtosis and the asymmetry of the observation error distribution.

2 Theoretical foundations

2.1 ML and M_P estimates

Let $P_{\theta_i} \in \mathcal{P}$ be distributions of independent random errors v_i and let they be indexed with the parameters: $E_i = E(v_i)$, $\mu_{2,i} = \mu_2(v_i) = \sigma_i^2$, $\mu_{3,i} = \mu_3(v_i)$ and $\mu_{4,i} = \mu_4(v_i)$. To estimate the moments μ_3 and μ_4 one can apply the methods proposed in the following papers (Wiśniewski 1995, 1996; Kasietczuk 1997). When we are interested in estimation of only one of these parameters, then it is usually assumed that the others are fixed (however, it is not neces-

sary to know their particular values). The set of distributions that are possible probabilistic models of geodetic observation errors will be denoted as $W_v = \{P_{\theta_i} : v_i \in \mathbb{S}_v, \theta_i \in \mathbb{S}_\theta \subset \Theta\}$; where \mathbb{S}_v is the set of possible values of random errors, for example, $\mathbb{S}_v = \langle a_1, a_2 \rangle$; \mathbb{S}_θ is the set of practically acceptable values of the parameter θ_i . Thus, we choose a probabilistic model of measurement errors from the distributions that belong to W_v (Wiśniewski 1987).

Let $\mathbf{y} = \mathbf{A}\mathbf{X} + \mathbf{v}$ be a conventional functional model of the observation vector $\mathbf{y} = [y_1, \dots, y_n]^T$ ($\mathbf{v} = [v_1, \dots, v_n]^T$ is a vector of random errors, $\mathbf{A} \in R^{n,r}$, $\text{rank}(\mathbf{A}) = r$, is a known matrix of coefficients, $\mathbf{X} \in (\Theta = R^r)$ is a vector of unknown parameters). Additionally, let independent observations $y_i = \mathbf{a}_i\mathbf{X} + v_i, i = 1, 2, \dots, n$, (\mathbf{a}_i – i th row of the matrix \mathbf{A}) have a fixed probability density function (PDF). Thus, each observation brings a piece of f -information $I(y_i; \mathbf{X}) = -\ln f(y_i; \mathbf{X}) = -\ln f(v_i)$ (see, e.g., Jones and Jones 2000; Wiśniewski 2009). Considering the parameter \mathbf{X} , all observations included within the vector \mathbf{y} bring the global amount of information as follows

$$\begin{aligned} I(\mathbf{y}; \mathbf{X}) &= \sum_{i=1}^n I(y_i; \mathbf{X}) = -\sum_{i=1}^n \ln f(y_i; \mathbf{X}) \\ &= -\sum_{i=1}^n \ln f(v_i) \end{aligned} \quad (2)$$

Note, that Eq. (2) is also an empirical version of Shannon's entropy (e.g. Ferrari and Yang 2010). Considering the global amount of information, one can search such estimate of the parameter \mathbf{X} which minimizes the amount of information included in the set of the observations (also information which is rather unexpected). Thus, the estimate $\hat{\mathbf{X}}$ should solve the optimization problem $\min_{\mathbf{X}} I(\mathbf{y}; \mathbf{X}) = I(\mathbf{y}; \hat{\mathbf{X}})$ (assuming that such extremum exists). For example, let $v_i \sim \text{ND}[E(v_i) = 0, \sigma_i^2]$ and $y_i \sim \text{ND}[E(y_i) = \mathbf{a}_i\mathbf{X}, \sigma_i^2]$, then on the basis of the following PDF, $f(y_i; \mathbf{X}) = f(v_i) \propto \exp(-p_i v_i^2/2)$, one can write $I(y_i; \mathbf{X}) = p_i v_i^2/2$ and hence $I(\mathbf{y}; \mathbf{X}) \propto \sum_{i=1}^n p_i v_i^2$. Since LS-estimate of the parameter \mathbf{X} minimizes the objective function $\phi(\mathbf{y}; \mathbf{X}) = \sum_{i=1}^n p_i v_i^2$ then, for the normal distributions, it also minimizes the amount of f -information related to the parameter \mathbf{X} . However, such estimates do not show such property for the other distribution which belongs to the set W_v . In such a case, the global amount of information is minimized by ML-estimates. Here, we look for the estimate $\hat{\mathbf{X}}$ which maximizes the likelihood

$$\mathcal{L}(\mathbf{y}; \mathbf{X}) = \ln \left(\prod_{i=1}^n f(v_i) \right) = \sum_{i=1}^n \ln f(v_i) \quad (3)$$

(see, e.g., Serfling 1980; Huber 1981; Koch 1990). Since $\max_{\mathbf{X}} \mathcal{L}(\mathbf{y}; \mathbf{X}) \Leftrightarrow \min_{\mathbf{X}} [-\mathcal{L}(\mathbf{y}; \mathbf{X}) = I(\mathbf{y}; \mathbf{X}) = \phi(\mathbf{y}; \mathbf{X})]$, where

$$\phi(\mathbf{y}; \mathbf{X}) = \sum_{i=1}^n [-\ln f(v_i)] = \sum_{i=1}^n \rho(v_i) \quad (4)$$

and $\rho(v_i) = -\ln f(v_i)$, then the idea of maximum likelihood is equivalent to the postulate of the minimum f -information related to $\hat{\mathbf{X}}$.

Now, let PDF be at least a twice differentiable function. Then, the Newton method can be applied to solve the optimization problem $\min_{\mathbf{X}} \phi(\mathbf{y}; \mathbf{X})$. Such solution uses the gradient and the Hessian of the function of Eq. (4) (see, e.g., Teunissen 1990). Thus, let us write the following form for the gradient

$$\begin{aligned} [\mathbf{g}_\phi(\mathbf{X})]^T &= \frac{\partial \phi(\mathbf{y}; \mathbf{X})}{\partial \mathbf{X}} = \frac{\partial}{\partial \mathbf{X}} \sum_{i=1}^n \rho(v_i) = \sum_{i=1}^n \frac{d\rho(v_i)}{dv_i} \frac{\partial v_i}{\partial \mathbf{X}} \\ &= -[\mathbf{b}(\mathbf{v})]^T \mathbf{A} \end{aligned} \quad (5)$$

where $\mathbf{b}(\mathbf{v}) = [b(v_1), \dots, b(v_n)]^T$ is a vector consisting of the following elements

$$\begin{aligned} b(v_i) &= \frac{d\rho(v_i)}{dv_i} = -\frac{d \ln f(v_i)}{dv_i} \\ &= -\frac{1}{f(v_i)} \frac{df(v_i)}{dv_i} = -\frac{f'(v_i)}{f(v_i)} \end{aligned} \quad (6)$$

and $\partial v_i / \partial \mathbf{X} = -\mathbf{a}_i$ (if it is convenient and not misleading, the derivatives $d^j f / d^j v_i, j = 1, 2, 3, 4, \dots$ will be denoted as $f', f'', f''', f^{(4)}, \dots$). If $v_i \sim \text{ND}[E(v_i) = 0, \sigma_i^2]$ and $f(v_i) \propto \exp(-p_i v_i^2/2)$, then $b(v_i) = p_i v_i$.

To find out the form of the Hessian of the objective function of Eq. (4), one can write

$$\mathbf{H}_\phi(\mathbf{X}) = \frac{\partial^2 \phi(\mathbf{X})}{\partial \mathbf{X} \partial \mathbf{X}^T} = \frac{\partial \mathbf{g}_\phi(\mathbf{X})}{\partial \mathbf{X}^T} = \mathbf{A}^T \frac{\partial \mathbf{b}(\mathbf{X})}{\partial \mathbf{X}^T} = \mathbf{A}^T \mathbf{R}(\mathbf{X}) \mathbf{A} \quad (7)$$

The diagonal elements of the matrix $\mathbf{R}(\mathbf{X}) = \text{Diag}[r(v_1), \dots, r(v_n)]$ are computed as the values of the following function

$$\begin{aligned} r(v_i) &= \frac{d^2 \rho(v_i)}{dv_i^2} = -\frac{d}{dv_i} \left(\frac{1}{f(v_i)} \frac{df(v_i)}{dv_i} \right) \\ &= -\frac{\frac{d^2 f(v_i)}{dv_i^2} f(v_i) - \left(\frac{df(v_i)}{dv_i} \right)^2}{[f(v_i)]^2} \end{aligned} \quad (8)$$

which is also known as the rigor function (e.g. Kamiński and Wiśniewski 1994). The necessary condition for the minimum of the function from Eq. (4) is that the Hessian from Eq. (7) be positively defined. It is obvious that such condition is satisfied for $\forall i : r(v_i) > 0$.

The gradient of Eq. (5) and the Hessian of Eq. (7) are the basis for the following iterative procedure ($j = 0, 1, \dots, k$)

$$\mathbf{X}^{j+1} = \mathbf{X}^j + d\mathbf{X}^{j+1}, \quad \mathbf{v}^{j+1} = \mathbf{y} - \mathbf{A}\mathbf{X}^{j+1} \quad (9)$$

with the iterative change of the parameter

$$\begin{aligned} d\mathbf{X}^{j+1} &= -[\mathbf{H}_\phi(\mathbf{X}^j)]^{-1} \mathbf{g}_\phi(\mathbf{X}^j) \\ &= [\mathbf{A}^T \mathbf{R}(\mathbf{X}^j) \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{b}(\mathbf{v}^j) \end{aligned} \quad (10)$$

The determination procedure of the ML-estimate of the vector \mathbf{X} ends for such $\hat{\mathbf{X}} = \mathbf{X}^k$ that $\mathbf{g}_\phi(\mathbf{X}^k) = \mathbf{0}$ and $(\mathbf{X}^{k+1} - \mathbf{X}^k = \mathbf{0}) \Leftrightarrow d\mathbf{X}^{k+1} = \mathbf{0}$.

ML-estimates can also be considered as a special case of the broader class of the estimates, namely M -estimates which are determined on the basis of influence functions. Let us consider a set of probability distributions $\mathcal{P} = \{P_{\mathbf{X}} : \mathbf{X} \in R^r\}$ with cumulative distributions $F(y_i; \mathbf{X}) = F(v_i)$, then one can write the related influence function in the following form (Hampel 1974; Serfling 1980; Huber 1981; Hampel et al. 1986; Yang 1997)

$$\text{IF}(y_i; \mathbf{X}, F) = -c_{\text{IF}} \psi(y_i, \mathbf{X}) = c_{\text{IF}} \psi(v_i) \frac{\partial v_i}{\partial \mathbf{X}^T} \quad (11)$$

where $c_{\text{IF}} = \int \psi' dF$. Due to the proportion $\text{IF}(y_i; \mathbf{X}, F) \propto \psi(v_i)$, the function $\psi(v_i)$ is also very often called the influence function. $\hat{\mathbf{X}}$ is an M -estimate of the parameter \mathbf{X} if it solves the following equation (see, e.g., Hampel et al. 1986; Yang 1999; Yang et al. 1999)

$$\begin{aligned} \sum_{i=1}^n \psi(y_i, \mathbf{X}) &= \sum_{i=1}^n \psi(v_i) \frac{\partial v_i}{\partial \mathbf{X}^T} \\ &= \psi(\mathbf{y}, \mathbf{X}) = \mathbf{A}^T \psi(\mathbf{v}) = \mathbf{0} \end{aligned} \quad (12)$$

where $\psi(\mathbf{v}) = [\psi(v_1), \dots, \psi(v_n)]^T$. The function $\psi(y_i, \mathbf{X})$ can be written (in relation to the objective function from Eq. (1)) as follows

$$\begin{aligned} \psi(y_i, \mathbf{X}) &= \frac{\partial \rho(y_i, \mathbf{X})}{\partial y_i} \frac{\partial y_i}{\partial \mathbf{X}^T} \\ &= \frac{d\rho(v_i)}{dv_i} \frac{\partial v_i}{\partial \mathbf{X}^T} = -\psi(v_i) \mathbf{a}_i^T \end{aligned} \quad (13)$$

where

$$\begin{aligned} \psi(v_i) &= \frac{d\rho(v_i)}{dv_i} = \frac{d\rho(v_i)}{d(v_i^2)} \cdot \frac{d(v_i^2)}{dv_i} \\ &= w(v_i) \frac{d(v_i^2)}{dv_i} = 2v_i w(v_i) \end{aligned} \quad (14)$$

Let

$$w(v_i) = \frac{d\rho(v_i)}{d(v_i^2)} = \frac{\psi(v_i)}{2v_i} \propto \frac{\psi(v_i)}{v_i} \quad (15)$$

be a weight function (Huber 1981; Hampel et al. 1986; Youcai and Mertikas 1995; Yang 1997). Then, for the fixed functions $\rho(v_i)$ one can write that $\mathbf{g}_\phi(\mathbf{X}) = \mathbf{A}^T \psi(\mathbf{v})$. Taking into account the expressions in Eqs. (14) and (15), and considering the following vector $\psi(\mathbf{v}) = \mathbf{w}(\mathbf{v})\mathbf{v}$, the Eq. (12) can be rewritten as $\mathbf{A}^T \psi(\mathbf{v}) = \mathbf{A}^T \mathbf{w}(\mathbf{v})\mathbf{v} = \mathbf{0}$, where $\mathbf{w}(\mathbf{v})$ is a diagonal matrix with the elements $\mathbf{w}(\mathbf{v})_{ii} = w(v_i)$. For the

model $\mathbf{v} = \mathbf{y} - \mathbf{A}\mathbf{X}$, one can obtain the following iterative solution

$$\hat{\mathbf{X}} = [\mathbf{A}^T \mathbf{w}(\hat{\mathbf{v}}) \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{w}(\hat{\mathbf{v}}) \mathbf{y} \quad (16)$$

where $\hat{\mathbf{v}} = \mathbf{y} - \mathbf{A}\hat{\mathbf{X}}$.

From a practical point of view, the approximation of the asymptotic covariance matrix is satisfactory in that case (Yang 1997). Thus, let

$$\mathbf{C}_{\hat{\mathbf{X}}} = \hat{\sigma}_{0,\psi}^2 [\mathbf{A}^T \mathbf{w}(\hat{\mathbf{v}}) \mathbf{A}]^{-1} \quad (17)$$

where $\hat{\sigma}_{0,\psi}^2$ is an iterative estimate of the variance coefficient computed as

$$\hat{\sigma}_{0,\psi}^2 = \frac{1}{n-r} \hat{\mathbf{v}}^T \mathbf{w}(\mathbf{v}) \hat{\mathbf{v}} \quad (18)$$

Note that Eqs. (17) and (18) follow from the application of the empirical influence function $\text{EIF}(y_i, \mathbf{X}) = [\mathbf{A}^T \mathbf{w}(\mathbf{v}) \mathbf{A}]^{-1} \mathbf{a}_i^T w(v_i) v_i$ (Yang 1997).

The estimate from Eq. (16), which solves the Eq. (12), also satisfies the necessary condition $\mathbf{g}_\phi(\mathbf{X}) = \mathbf{0}$ for the minimum of the function from Eq. (1). For $\rho(v_i) = -\ln f(v_i)$, the following equality holds (see, e.g., Youcai and Mertikas 1995; Dzhun' 2011)

$$w(v_i) = \frac{d\rho(v_i)}{d(v_i^2)} \propto \frac{d\rho(v_i)}{v_i dv_i} = -\frac{1}{v_i f(v_i)} \frac{df(v_i)}{dv_i} \quad (19)$$

The estimate from Eq. (16) together with the weight function from Eq. (19) can be regarded as $M_{\mathcal{P}}$ estimator with the explicit probabilistic model $\mathcal{P} = \{P_{\mathbf{X}} : \mathbf{X} \in R^r\}$.

In the case of LS-method, where $\rho(v_i) = p_i v_i^2$, we get $\psi(v_i) = 2p_i v_i$ and $w(v_i) = p_i \propto \psi(v_i)/v_i$. Such variant of $M_{\mathcal{P}}$ estimation is related to the set of normal distributions $\mathcal{P} = \{\text{ND}[E(v_i) = 0, \sigma_i^2]\}$. Another example of the probabilistic models in $M_{\mathcal{P}}$ estimation is the following set $\mathcal{P} = \{(1-\varepsilon)P_{\mathbf{X},\alpha} + \varepsilon P_{\mathbf{X},\beta} : \mathbf{X} \in R^r\}$, $0 \leq \varepsilon \leq 1$, which is the basis for robust M -estimation (Huber 1981; Hampel et al. 1986). Here, we assume that $P_{\mathbf{X},\alpha} \in W_v$ is an acceptable distribution of measurement errors, while $P_{\mathbf{X},\beta}$ is an unacceptable (strange) distribution. Considering LS-method with the equivalent weights $\hat{p}_i \propto \hat{\mu}_{2,i}^{-1}$, we can write that $P_{\mathbf{X},\alpha} = P_{\mathbf{X}/\mu_2} = \text{ND}[\mathbf{a}_i \mathbf{X}, \sigma_i^2]$ and $P_{\mathbf{X},\beta} = P_{\mathbf{X}/\hat{\mu}_2} = \text{ND}[\mathbf{a}_i \mathbf{X}, \hat{\sigma}_i^2]$, where $\hat{\sigma}_i^2 = \hat{\mu}_{2,i}$ is the equivalent variance.

Remember that usually $\hat{p}_i = t(v_i)p_i$, where $0 \leq t(v_i) \leq 1$ is an attenuation function (e.g., Krarup and Kubik 1983; Yang 1994; Yang et al. 2002). In such a case, the equivalent variance can be written as $\hat{\mu}_{2,i} = \mu_{2,i}/t(v_i)$. Probabilistic models are also applied in a development of M -estimation, namely M_{split} estimation (Wiśniewski 2009, 2010; Duchnowski and Wiśniewski 2011). In the basic variant of the method in question, it is assumed that $\mathcal{P} = \{P_{\mathbf{X}_\alpha}, P_{\mathbf{X}_\beta} : \mathbf{X}_\alpha, \mathbf{X}_\beta \in R^r\} \subset W_v$. \mathbf{X}_α and \mathbf{X}_β are acceptable variants

of the same parameter \mathbf{X} . Assuming that \mathcal{P} is a set of normal distributions that differ from each other in the parameter \mathbf{X} , we get the squared M_{split} estimation with the following components of the objective function $\rho(\mathbf{X}_\alpha, \mathbf{X}_\beta) = v_{i\alpha}^2 v_{i\beta}^2$.

2.2 Influence, weight and rigor functions in $M_{\mathcal{P}}$ estimation: robustness of the method

When \mathcal{P} is a set of distributions which are symmetric about $v=0$, then $E=E(v)=0$ and the mode M_0 is at the origin of the coordinate system, namely $M_0=0$. For such distributions, it holds that $\psi(v)=-\psi(-v)$, and the weight functions $w(v)$ as well as the rigor functions are symmetric ones. For asymmetric distributions, the expected value moves away from the mode with the increase of the asymmetry coefficient, with $\text{sgn}(\delta) = \text{sgn}(\gamma_1)$, where $\delta = E - M_0$ is the shift between the expected value and the mode. For such distributions the origin on the coordinate system might lay: except the mode ($\text{PDF}_{(M_0 \neq 0)}$), at the mode ($\text{PDF}_{(M_0=0)}$) or at the expected value ($\text{PDF}_{(E=0)}$).

For the mode of the variable v and for the arbitrary $\epsilon > 0$ it holds that $\forall v_0 \neq M_0 : P(v \in \Delta M_{0(\epsilon)}) > P(v \in \Delta v_{0(\epsilon)})$ where $\Delta M_{0(\epsilon)} = \langle M_0 - \epsilon, M_0 + \epsilon \rangle$ and $\Delta v_{0(\epsilon)} = \langle v_0 - \epsilon, v_0 + \epsilon \rangle$. The postulate of the minimum information is satisfied by the application of the individual information $I(y; \mathbf{X}) = I(v - s) = I(\tilde{v})$ where $s = M_0$ for $\text{PDF}_{(M_0 \neq 0)}$, $s = 0$ for $\text{PDF}_{(M_0=0)}$ and $s = -\delta$ for $\text{PDF}_{(E=0)}$. In such case, $M_{\mathcal{P}}$ estimates should follow from the influence function determined on the basis of the objective function which is shifted, namely on the basis of its components $\rho(v - s) = \rho(\tilde{v})$. Thus

$$\psi_s(v) = \psi(\tilde{v}) = \frac{d\rho(v - s)}{dv} = \frac{d\rho(\tilde{v})}{d(v^2)} \frac{d(v^2)}{dv} = 2vw_s(v) \quad (20)$$

where

$$w_s(v) = w(\tilde{v}) = \frac{d\rho(\tilde{v}_s)}{d(v^2)} = \frac{\psi_s(v)}{2v} \propto \frac{\psi_s(v)}{v} \quad (21)$$

is a weight function which is determined with respect to the mode M_0 . If $v = M_0 \neq 0$ then $w_s(M_0) = \psi_s(M_0)/M_0$. However, if $M_0 = 0$ then $\lim_{v \rightarrow 0} w_s(v) = \psi'_s(0) = r_s(0)$ (such value was derived by Kadaj (1988) for the symmetric distributions and for the conventional influence function). Considering the functions from Eqs. (20) and (21), one can obtain the following rigor function

$$\begin{aligned} r_s(v) = r(\tilde{v}) &= \frac{d^2\rho(\tilde{v})}{dv^2} \\ &= \frac{d\psi_s(v)}{dv} = 2[w_s(v) + vw'_s(v)] \end{aligned} \quad (22)$$

If the weight function $w_s(v)$ has a maximum at $v = M_0$, then $w'_s(M_0) = 0$ and $r_s(M_0) = 2w_s(M_0)$. It does not follow that

every rigor function of Eq. (22) has a maximum at that point. Considering that $r'_s(v) = 2[2w'_s(v) + vw''_s(v)]$ it holds only if $w''_s(M_0) = 0$. From a theoretical point of view, such property is adverse; however, one can propose a following corrected rigor function

$$\begin{aligned} r_{s,M_0}(v) &= 2[w_s(v) + (v - M_0)w'_s(v)] \\ &= r_s(v) - 2M_0w'_s(v) \end{aligned} \quad (23)$$

that is free of such property. Note that for such corrected rigor function $r'_{s,M_0}(v) = 4w'_s(v) + 2vw''_s(v) - 2M_0w''_s(v)$, hence $\max_v r_{s,M_0}(v) = r_{s,M_0}(M_0)$ holds independently of the value of the derivative $w''_s(v)$ at M_0 .

While analyzing the expected properties of $M_{\mathcal{P}}$ estimates we can apply the influence function $\text{IF}(y; \mathbf{X}, F) \propto \psi(v)$ to determine the following parameters: the maximum sensitivity to gross errors, $\gamma^* = \sup_v |\psi(v)|$; the local-shift sensitivity, $\lambda^* = \sup_v |[\psi(v + \delta v) - \psi(v)]/\delta v|$ and the rejection point $\rho_v^* = \inf \{v^* : \psi(v) = 0, |v| > v^*\}$ (Hampel 1974; Hampel et al. 1986). In the case of asymmetric distributions, one can consider left- or right-handed rejection points, namely $\rho_v^* = \inf \{v_-^* : \psi(v) = 0, v < v_-^*\}$ and $\rho_v^+ = \inf \{v_+^* : \psi(v) = 0, v > v_+^*\}$. For $\gamma_1 = 0$

we obtain $|\rho_v^*| = |\rho_v^+| = \rho^*$. Additionally, we also consider a parameter which describes the maximum sensitivity to observation disturbances

$$\begin{aligned} \lambda_r^* &= \sup_v \lim_{\delta v \rightarrow 0} \left| \frac{\psi_s(v + \delta v) - \psi_s(v)}{\delta v} \right| \equiv \sup_v \left| \frac{d\psi_s(v)}{dv} \right| \\ &= \sup_v |r_s(v)| \end{aligned} \quad (24)$$

Such parameter is of course related to the sensitivity λ^* (for the influence function $\text{IF}(y - s; \mathbf{X}, F) \propto \psi(v - s) = \psi_s(v)$). The important problem is that this parameter is not achieved at the mode. However, if it relates to the corrected rigor function, namely if

$$\lambda_{r,M_0}^* = \sup_v |r_{s,M_0}(v)| \quad (25)$$

then $\max_v r_{s,M_0}(v) = r_{s,M_0}(M_0) = \lambda_{r,M_0}^* \Leftrightarrow \max_v w_s(v) = w_s(M_0) = \lambda_w^*$ (if only such extreme points exist).

The class of M -estimates can be divided into the following subclasses (Kadaj 1988): robust K_- , neutral K_0 and weak K_+ . Let us extend such classification to $M_{\mathcal{P}}$ estimates, thus let

$$\begin{aligned} M_{\mathcal{P}} \in K_- &\text{ when } \forall v_i \neq \Delta M_{0(l)}, v_j \notin \Delta M_{0(l)} : \\ &w_s(v_i) \leq \lambda_w^*, w_s(v_j) < \lambda_w^* \\ M_{\mathcal{P}} \in K_0 &\text{ when } \forall v : w_s(v) = \lambda_w^* \\ M_{\mathcal{P}} \in K_+ &\text{ when } \forall v_i \neq \Delta M_{0(l)}, v_j \notin \Delta M_{0(l)} : \\ &w_s(v_i) \geq \lambda_w^*, w_s(v_j) > \lambda_w^* \end{aligned} \quad (26)$$

where $\Delta M_{0(l)} = \langle M_0 - l, M_0 + l \rangle, l \geq 0$. For example, in the case of LS-method and the set of the normal distributions we have $\forall v : w_s(v) = \lambda_w^* = p$, hence

LS $\equiv M_{\mathcal{P}=\{\text{ND}[0,\sigma^2]\}} \in K_0$. In the case of M -estimation with the probabilistic model $\mathcal{P} = \{(1-\varepsilon)P_\alpha + \varepsilon P_\beta\}$ with implicit distributions P_α and P_β , and with application of the Huber function (Huber 1981), one can write the following forms of the weight functions (for $\bar{v} = v/\sigma$ and, e.g., $l = 2.0, 2.5, 3.0$)

$$w(\bar{v}) = \begin{cases} 1 & \text{for } |\bar{v}| \leq l \\ \frac{l}{|\bar{v}|} & \text{for } |\bar{v}| > l \end{cases} \quad (27)$$

Here $\forall \bar{v} \in \Delta \bar{v} = \langle -l, l \rangle : w(\bar{v}) = \lambda_w^*$ and $\forall \bar{v} \notin \Delta \bar{v} : w(\bar{v}) < \lambda_w^*$, where $\lambda_w^* = 1$ ($\Delta \bar{v} = \langle -l, l \rangle$ is the interval for acceptable standardized random errors \bar{v}). Thus, such method of M -estimation, which will be denoted as M_H , belongs to the subclass K_- . On the other hand $M_{\mathcal{P}}$ estimations which have U-shaped weight functions will belong to the subclass of weak estimations K_+ .

3 Probabilistic models

3.1 Pearson distributions

The Pearson distribution system can play an important role in $M_{\mathcal{P}}$ estimation with the weight function from Eq. (19). The probability density functions of PD-system are solutions of the following differential equation (Pearson 1920; Elderton 1953; Hald 2007; Nagahara 2007, 2011; Friori and Zenga 2009; Dzhun' 2011; Xi et al. 2012)

$$\begin{aligned} \Omega(v) &= \frac{1}{f(v)} \frac{df(v)}{dv} = \frac{d \ln f(v)}{dv} \\ &= -\frac{(c_0 + 3c_2)v - \sigma c_1}{\sigma^2 c_0 - \sigma c_1 v + c_2 v^2} \end{aligned} \quad (28)$$

where:

$$c_0 = 4\beta_2 - 3\beta_1, \quad c_1 = \gamma_1(\beta_2 + 3), \quad c_2 = 2\beta_2 - 3\beta_1 - 6 \quad (29)$$

The papers cited above also showed other variants of the differential equation in question which depends on the parameter notation. The special role of the Pearson distributions in M -estimation with probabilistic models follows from the fact that Eq. (28) can be regarded as the influence function of $M_{\mathcal{P}}$ estimation, namely $\psi(v) = -\Omega(v) = -f'(v)/f(v)$. Such assumption refers to the general proposition of the application the expression $\Omega(v)/v$ as the weight function in ML-method (Dzhun' 2011).

The asymmetry $\beta_1 = \gamma_1^2$ and kurtosis β_2 are the parameters in Eq. (28), hence they are also parameters of the influence function $\psi(v) = -\Omega(v)$. This allows us to obtain M -estimates which can consider anomalies in observation distributions (in this paper context). The range of the steering parameters of M -estimation might have also another meaning. For example, suppose that β_2 is not necessarily the real

kurtosis of the distribution, then it can be the parameter which influences the robustness of M -estimates (this will be noted in the next part of the paper).

The Eq. (28) has three basic solutions, namely three function types: type I, IV and VI and additionally 10 transitional functions (including the probability density function of the normal distribution). The solution in question depends on the roots of the equation $\sigma^2 c_0 - \sigma c_1 v + c_2 v^2 = 0$, and are classified by the criterion (e.g. Elderton 1953; Hald 2007)

$$k = \frac{c_1^2}{4c_0c_2} = \frac{\beta_1(\beta_2 + 3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)} \quad (30)$$

If $k < 0$, then two real roots exist which differ from each other in the sign. In such a case we get PDF of type I (PDF*PI $_{\beta_1>0, \beta_2<3}$), which under $\beta_1 = 0$ changes into type II (PDF*PII $_{\beta_1=0, \beta_2<3}$), and if additionally $\beta_2 = 3$ it changes into PDF of the normal distribution (PDF*ND $_{\beta_1=0, \beta_2=3}$). PDFs of the types I and II have limited range and are usually bell-shaped, but may be U-shaped and J-shaped (Elderton 1953). If $0 < k < 1$, then the roots are complex, and the solution of the Eq. (28) is the function of type IV, for which

$$\begin{aligned} \text{PDF*PIV}_{\beta_1>0, \beta_2>3} &\xrightarrow{\beta_1=0} \text{PDF*PVII}_{\beta_1=0, \beta_2>3} \\ &\xrightarrow{\beta_2=3} \text{PDF*ND}_{\beta_1=0, \beta_2=3} \end{aligned} \quad (31)$$

Probability density functions from Eq. (31) have unlimited range and are bell-shaped. If $k > 1$, the roots are real and of the same sign, and one can get the curve of type VI (if $k = 1$, then we can get the curve of type V). Those curves as well as other transitional curves are rarely applied for PDF. Since distributions of geodetic observations are usually leptokurtic, most of all we will be interested in the distributions PIV ($0 < k < 1$) and PVII ($k = 0$) (note the general relation from Eq. (31)). The leptokurtic model which includes the chosen classes of the distributions will be denoted as $\mathcal{P}_{\text{PD}(l)} = \{\text{PIV}, \text{PVII} : \sigma, \beta_1, \beta_2 > 3\}$. $M_{\mathcal{P}}$ estimation which applies such a model will be denoted as $M_{\text{PD}(l)}$. Consequently, $\mathcal{P}_{\text{PD}(p)} = \{\text{PI}, \text{PII} : \sigma, \beta_1, \beta_2 < 3\}$ is a model of platykurtic distributions and $\mathcal{P}_{\text{PD}(m)} = \{\text{PVII} : \sigma, \beta_1, \beta_2 = 3\}$ is a model of mesokurtic distributions.

Considering the model $\mathcal{P}_{\text{PD}(l)}$, then k should be within the interval $0 \leq k < 1$. If $k = 0$ and additionally $\beta_1 = 0, \beta_2 = 3$, then $\mathcal{P}_{\text{PD}(l)} \rightarrow \mathcal{P}_{\text{PD}(m)} \equiv \mathcal{P}_{\text{ND}}$. The conditions concerning the value of k entail limitations of values of β_1 and β_2 . Generally speaking, in the case of PDF*PIV $_{\beta_1>0, \beta_2>3}$ the tolerance of increasing asymmetry of the platykurtic distribution becomes greater in the case of increasing kurtosis. If the excess is small and the asymmetry is large, then for such PDF k exceeds the acceptable limits (Fig. 1).

For the function (28), the origin of the coordinate system is at the point $E = E(v) = 0$ (Elderton 1953). The distance between the expected value $E = \sigma c_1/2c_2$ and the mode $M_0 = \sigma c_1(c_0 + c_2)/2c_2(c_0 + 3c_2)$ is equal to $\delta = E - M_0 =$

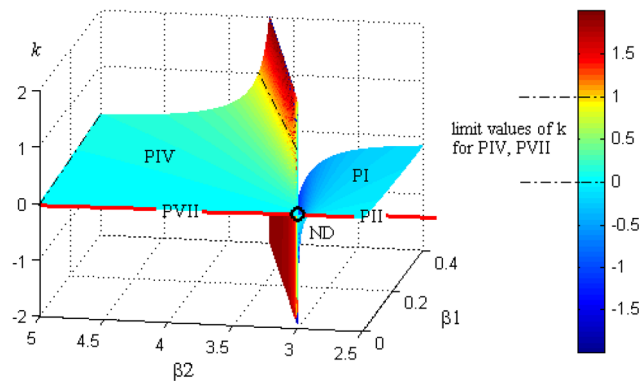


Fig. 1 Values of the criterion k

$\sigma c_1/(c_0 + 3c_2)$ (for PVII $\beta_1 = 0$, hence $c_1 = 0$ and $\delta = 0$). Thus creating the shifted influence function $\psi(v-s) = \psi(\tilde{v})$ (with the shift $s = -\delta$) one can write

$$\begin{aligned}\psi(\tilde{v}) &= \psi_\delta(v) \\ &= \frac{(c_0 + 3c_2)(v + \delta) - \sigma c_1}{\sigma^2 c_0 - \sigma c_1(v + \delta) + c_2(v + \delta)^2} \\ &= \frac{(c_0 + 3c_2)v}{\sigma^2 c_0 - \sigma c_1 \tilde{v} + c_2 \tilde{v}^2}\end{aligned}\quad (32)$$

Such influence function is related to the following PDF'PIV, which mode M_0 coincides with the origin (Elderton 1953; Wiśniewski 1987, 1989)

$$f(v) \propto (1 + v_{M_0,a}^2)^{-m} \exp(-\vartheta \arctg v_{M_0,a}^2) \quad (33)$$

where: $v_{M_0,a} = (v - M_0)/a$, $a = \sigma z/4$, $\vartheta = \gamma_1(c_0^2 - c_2^2)/z c_2^2$, $m = (c_0 + 3c_2)/2c_2$, $z = 2(4c_0 c_2 - c_1^2)^{1/2}/c_2$. For such PDF one can write the objective function of ML-method as follows $\rho(v) = m \ln(1 + v_{M_0,a}^2) + \vartheta \arctg v_{M_0,a}^2$. If $\beta_1 = 0$ and $v_{M_0,a} = v_a = v/a$, then one can obtain PDF'PVII in the form $f(v) \propto (1 + v_a^2)^{-m}$, and hence $\rho(v) = m \ln(1 + v_a^2)$.

The weight function which corresponds to the influence function from Eq. (32) has the following form

$$w_\delta(v) = \frac{\psi_\delta(v)}{v} = \frac{(c_0 + 3c_2)}{\sigma^2 c_0 - \sigma c_1 \tilde{v} + c_2 \tilde{v}^2} \quad (34)$$

Now let us find the point at which such function has the maximum. The derivative $w'_\delta(v)$ is equal to zero for such $\tilde{v} = v + \delta = v + \sigma c_1/(c_0 + 3c_2)$, for which

$$-\sigma c_1(c_0 + 3c_2) + 2c_2(c_0 + 3c_2)v + 2\sigma c_2 c_1 = 0 \quad (35)$$

and hence

$$v = \frac{\sigma c_1(c_0 + 3c_2) - 2\sigma c_2 c_1}{2c_2(c_0 + 3c_2)} = \frac{\sigma c_1(c_0 + c_2)}{2c_2(c_0 + 3c_2)} = M_0 \quad (36)$$

If $\beta_2 > 3$, then for each $v \neq M_0$ it holds that $r_{\delta,M_0}(v) \leq \lambda_{r,M_0}^*$ and $w_\delta(v) < \lambda_w^*$. Thus, M_P estimates, in which the model $\mathcal{P}_{PD(I)} = \{\text{PIV, PVII} : \sigma, \beta_1 \geq 0, \beta_2 > 3\}$ is applied, belong to the class of robust estimation, namely $M_{PD(I)} \in K_-$. The

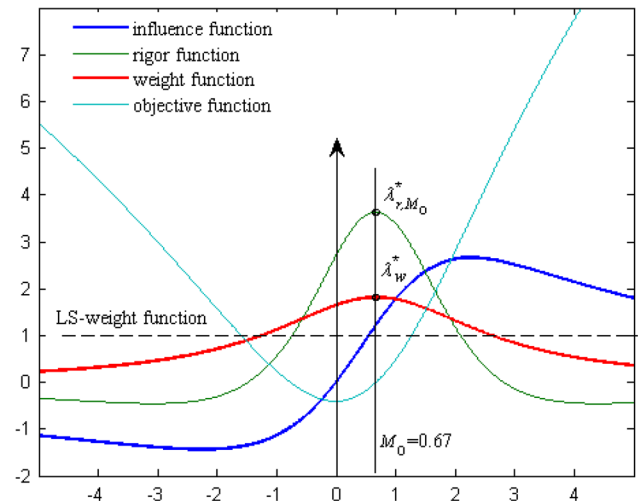


Fig. 2 Functions characterizing ML and M_P estimations; for $\sigma = 1$, $\beta_1 = 0.64$, $\beta_2 = 6.00$ ($k = 0.14$, $c_0 = 22.08$, $c_1 = 7.20$, $c_2 = 4.08$)

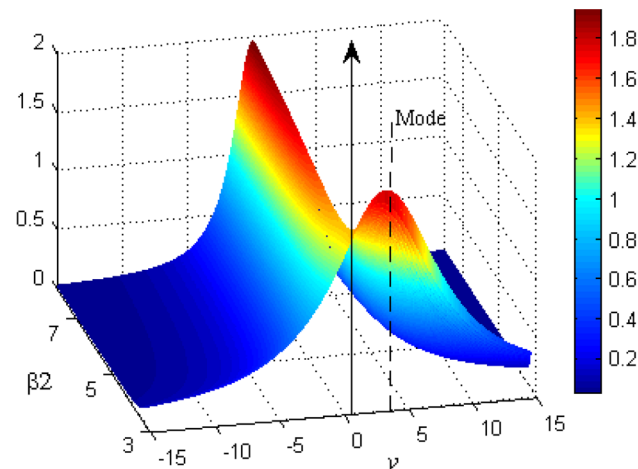


Fig. 3 Weight functions of M_P estimation for $\mathcal{P}_{PD(I)} = \{\text{PIV} : \sigma = 1, \beta_1 = 0.16, 3.4 < \beta_2 < 8\}$

graphs of the functions which characterize that method are presented in Fig. 2. The graphs of the weight functions for the models $\mathcal{P}_{PD(I)} = \{\text{PIV} : \sigma = 1, \beta_1 > 0, \beta_2 > 3\}$ and $\mathcal{P}_{PD(I)} = \{\text{PVII} : \sigma = 1, \beta_1 = 0, \beta_2 > 3\}$ are presented in Figs. 3 and 4.

For the sake of comparing, let us now consider the weight functions of M_P estimation with the platykurtic Pearson distributions of type I and II. If the mode is at the origin, then PDF'PI can be written as follows (Elderton 1953; Wiśniewski 1987, 1989)

$$f(v) \propto (1 + v/a_1)^{m_1} (1 - v/a_2)^{m_2} \quad (37)$$

where: $m_1 = -2m + \vartheta$, $m_2 = -2m - \vartheta$. The quantity z , which appears in the parameter ϑ (see, Eq. (33)), can be written as $z = 2(c_1^2 - 4c_0 c_2)^{1/2}/c_2$. For $\beta_1 = 0$ one can write that $f(v) \propto (1 - v^2/a^2)^m$. For the distributions PI and PII, the variable v can vary within the interval $\langle a_1, a_2 \rangle$, with

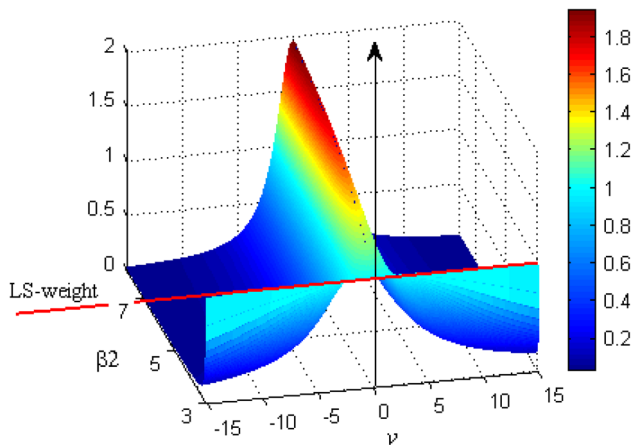


Fig. 4 Weight functions of $M_{\mathcal{P}}$ estimation and LS for $\mathcal{P}_{PD(l)} = \{\text{PVII}: \sigma = 1, \beta_1 = 0, 3 < \beta_2 \leq 8\} \rightarrow \beta_2 \rightarrow 3 + \mathcal{P}_{ND} = \{\text{ND}: \sigma = 1, \beta_1 = 0, \beta_2 = 3\}$

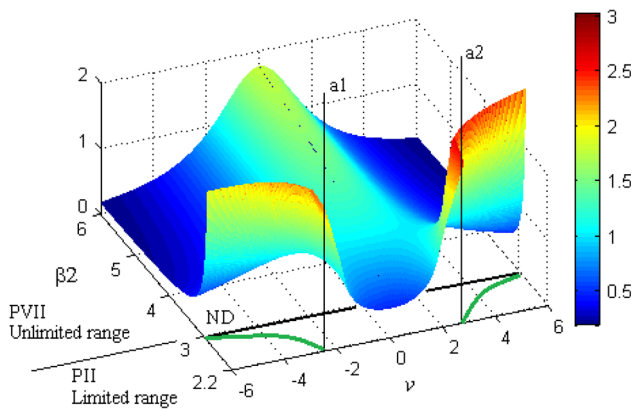


Fig. 5 Weight function of $M_{\mathcal{P}}$ estimation for the mixed model $\mathcal{P}_{PD(l,p)} = \{\text{PII}, \text{PVII}: \sigma = 1, \beta_1 = 0, 2.2 \leq \beta_2 \leq 6\}$. If $\beta_2 = 2.2$, then $a_1 = -2.16, a_2 = 2.16$

$a_1 = \sigma m_1 z / 2(m_1 + m_2)$ and $a_2 = -a_1 m_2 / m_1$. If $\beta_1 = 0$, then $c_0 = 4\beta_2, c_1 = 0, c_2 = 2\beta_2 - 6, \vartheta = 0, m_1 = m_2 = m$ and $z = 2(-4c_0 c_2)^{1/2} / c_2$ (for $\beta_2 < 3$ and acceptable values of the asymmetry coefficient β_1 , the quantity c_2 has negative values). In such a case $a_1 = \sigma m z / 4m = \sigma z / 4$, then for the distribution PII we can write $a_1 = -a_2 = -[2\beta_2 / (\beta_2 - 3)]^{1/2}$.

The weight functions of $M_{\mathcal{P}}$ estimation with symmetric leptokurtic and platykurtic distributions $\mathcal{P}_{PD(l,p)} = \{\text{PII}, \text{PVII}: \sigma = 1, \beta_1 = 0, \beta_2\}$ is presented in Fig. 5.

For the distributions PI and PII it holds that $\forall v \neq M_0 : w_8(v) > \lambda_w^*$. Thus, $M_{\mathcal{P}}$ estimates for $\mathcal{P}_{PD(p)} = \{\text{PI}, \text{PII}: \sigma, \beta_1 \geq 0, \beta_2 < 3\}$ belong to the class of the weak estimates K_+ . Such models might be interesting in some particular estimation problems. However, such problems are beyond the scope and objective of this paper, thus the models in question will not be discussed here anymore.

3.2 Gram–Charlier series

Assume that the set of distributions $\mathcal{P} = \{f(v) : \mu_k, k = 2, 3, 4, \dots\} \subset W_v$ is not much different from the set $\mathcal{P}_{PD(m)} = \{\text{ND} : \sigma, \gamma_1 = 0, \gamma_2 = 0\}$. Then, the function $f(v)$ can be replaced by the sum $f(v) = f_{\text{ND}}(v) + \zeta(v; \mu_2, \mu_3, \dots)$, where the remainder term $\zeta(v)$ is expanded as a series of v (e.g. Elderton 1953). Here, we will apply an orthogonal expansion, namely the Gram–Charlier A series (Charlier 1906; Tomozawa 1974; Berberan-Santos 2007). A similar asymptotic expansion was also proposed by Edgeworth (see, e.g., Elderton 1953; Crawford and Walsh 1960; Kolassa 2006). The Gram–Charlier series (G–C) can be presented as polynomials which are orthogonal with respect to PDF'ND, namely (e.g. Hald 2007)

$$\begin{aligned} f(v) &= \tilde{f}(v)_{\infty} = \sum_{j=0}^{\infty} \lambda_j \frac{\sigma^j}{j!} \frac{d^j f_{\text{ND}}(v)}{dv^j} \\ &= \sum_{j=0}^{\infty} \lambda_j \frac{\sigma^j}{j!} (-1)^j \sigma^{-j} H_j(\bar{v}) f_{\text{ND}}(v) \\ &= f_{\text{ND}}(v) h(\bar{v})_{\infty} \propto \exp(-v^2/2\sigma^2) h(\bar{v})_{\infty} \end{aligned} \quad (38)$$

where $h(\bar{v})_{\infty} = \sum_{j=0}^{\infty} \lambda_j (-1)^j H_j(\bar{v}) / j!$ (for $\bar{v} = v/\sigma$) in which, $H_j(\bar{v})$ is the Hermite polynomial of degree j for which $H_j^{(i)}(\bar{v}) = j H_{j-1}^{(i-1)}(\bar{v})$. We usually apply the first few polynomials of G–C series. For example, if $j = 4$ then: $H_0(\bar{v}) = 1, H_1(\bar{v}) = \bar{v}, H_2(\bar{v}) = \bar{v}^2 - 1, H_3(\bar{v}) = \bar{v}^3 - 3\bar{v}, H_4(\bar{v}) = \bar{v}^4 - 6\bar{v}^2 + 3$.

If PDF in the set of distributions $\mathcal{P} = \{f(v) : \sigma, \gamma_1, \gamma_2\}$ is replaced by the function from Eq. (38), then the component of the objective function of ML-method with the approximating model $\tilde{\mathcal{P}}_{G-C} = \{\tilde{f}_{\infty}(v) : \sigma, \gamma_1, \gamma_2\}$ takes the following form

$$\rho(v) = -\ln \tilde{f}_{\infty}(v) = \frac{v^2}{2\sigma^2} - \ln h(\bar{v})_{\infty} \quad (39)$$

This is the basis for determination of the influence function of $M_{\mathcal{P}}$ estimation as well as the rigor function of ML-method, hence

$$\psi(v) = \frac{d\rho(v)}{dv} = \frac{v}{\sigma^2} - \frac{h'(\bar{v})}{\sigma h(\bar{v})} \quad (40)$$

$$r(v) = \frac{d^2\rho(v)}{dv^2} = \frac{1}{\sigma^2} - \frac{h''(\bar{v})h(\bar{v}) - [h'(\bar{v})]^2}{\sigma^2[h(\bar{v})]^2} \quad (41)$$

The orthogonality condition can be used to determine the coefficients λ_i , thus (see, e.g., Elderton 1953): $\lambda_0 = 1, \lambda_1 = \lambda_2 = 0, \lambda_3 = -\gamma_1, \lambda_4 = \gamma_2, \lambda_5 = -\mu_5/\sigma^5 + 10\mu_3/\sigma^3$ etc. The terms of the G–C series for which $j > 4$ include the respective moments of the order higher than $k = 4$. In accordance with the paper assumptions (the distribution parameters are the moments of at most the fourth order), we will

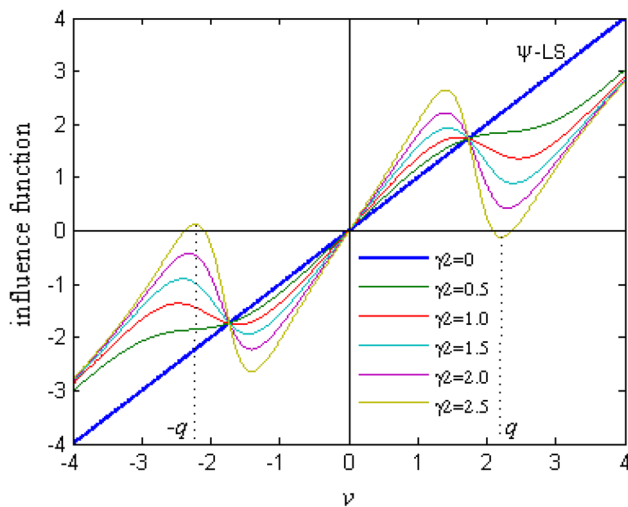


Fig. 6 Influence function for M_P estimation with the approximating model $\tilde{\mathcal{P}}_{G-C} = \{\tilde{f}(v)_4 : \sigma = 1, \gamma_1 = 0, \gamma_2\}$

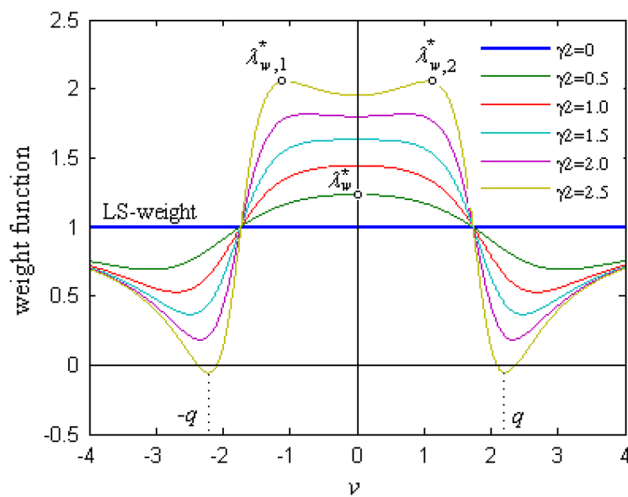


Fig. 7 Weight function for M_P estimation with the approximating model $\tilde{\mathcal{P}}_{G-C} = \{\tilde{f}(v)_4 : \sigma, \gamma_1 = 0, \gamma_2\}$

not consider such terms of the G–C series. Such limitation of the expansion leads to the following expression

$$h(\bar{v})_4 = 1 + \frac{1}{6}\gamma_1 H_3(\bar{v}) + \frac{1}{24}\gamma_2 H_4(\bar{v}) \quad (42)$$

and to the respective derivatives. Considering estimation of the parameters from the functional model $\mathbf{y} = \mathbf{A}\mathbf{X} + \mathbf{v}$, the series $\tilde{f}(v)_4 = f_{ND}(v)h(\bar{v})_4$ can be sufficient in many cases, especially for small asymmetry and moderate excess. However, the G–C series that are discussed here are not expected to have the same properties as non-approximated PDFs (for example, for particular intervals of the variable v , $\tilde{f}(v)_4$ might attain negative values). The functions that characterize M_P estimation with the approximating model $\tilde{\mathcal{P}} = \{\tilde{f}(v)_4 : \sigma, \gamma_1, \gamma_2\}$, which will be denoted as M_{G-C} , are presented in Figs. 6 and 7.

The influence and weight functions show some unfavorable features for bigger values of observation errors and growing excess (e.g., there is a point q from which the functions tend to the respective functions of LS-method). For some values of the excess, there are also points $\lambda_{w,1}^*$ and $\lambda_{w,2}^*$ at which the weight function has the local maxima (see, Fig. 7). When analyzing the weight function $w(v) = \psi(v)/v$ which is related to the influence function (40), it is worth noting that for the non-zero excess, $\lim_{v \rightarrow 0^-} w(v) = \lambda_{w,1}^* = -\infty$ and $\lim_{v \rightarrow 0^+} w(v) = \lambda_{w,2}^* = +\infty$ (fluctuation of the weight function). Thus, in the close neighborhood of the point $v = 0$, the weight function should be numerically corrected. For example, let $|v_i| < \epsilon_v$ then $w(v_i) := w(v_{i-1})$, where ϵ_v is the boundary point for acceptable fluctuation of the weight function. Note, that such discontinuity does not concern the rigor function $r(v)$ (in the next section, M_P estimates with the correction of the weight function are verified by the corresponding ML-estimates)

4 Numerical tests

The basic properties of M_P estimation with the probabilistic model $\mathcal{P}_{PD(I)} = \{\text{PIV, PVII} : \sigma, \beta_1, \beta_2 > 3\}$ ($M_{PD(I)}$ method) as well as with the approximating model $\tilde{\mathcal{P}}_{G-C} = \{\tilde{f}(v)_4\}$ (M_{G-C} method), are illustrated with an example of a simulated levelling network (Fig. 8). The network consists of the five fixed points and four new points (A, B, C, D) and 16 height differences which are measured. It is assumed that each height difference is measured s times (in the main tests $s = 4$). Thus, the known matrix of coefficient can be written as $\mathbf{A} = \mathbf{A}_0 \otimes \mathbf{1}_s$, where $\mathbf{A}_0 \in R^{n,r}$ is a matrix related to single measurements of the height differences

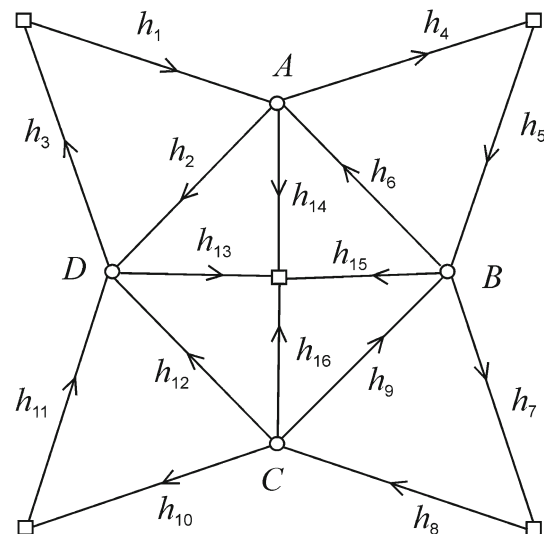


Fig. 8 Simulated levelling network

h_i , and $\mathbf{1}_s = [1, \dots, 1_s]^T$ (\otimes is the Kronecker product). All point heights are assumed to be equal to zero, namely $\mathbf{X} = [H_A, H_B, H_C, H_D]^T = \mathbf{0}$, hence all measurement results are equal to respective simulated measurement errors.

In accordance with the objective of the paper, the measurement errors are characterized by non-zero asymmetry and the excess. For small samples (for example, when $n = 64$), sets which are simulated by the Gaussian random number generators show some anomalies in comparison with the assumed normal distribution. Such property concerns also the one-dimensional generator $\text{randn}(n, 1)$ (n is a number of the generated numbers) which is the component of MatLab. We are interested in anomalies which concern the asymmetry and the excess of measurement distribution. Thus, the simulation is to generate numbers $v_i, i = 1, \dots, n$, with the application of the function in question (for the theoretical values of $E(v) = 0, \sigma = 1$ and $\gamma_1 = 0, \beta_2 = 3$). Then, for such generated set, one can compute the empirical moments $\bar{E} = n^{-1} \sum_{i=1}^n v_i, \bar{\mu}_k = n^{-1} \sum_{i=1}^n (v_i - \bar{E})^k, k = 2, 3, 4$, and hence the empirical values of the asymmetry coefficients $\bar{\gamma}_1 = \bar{\mu}_3/(\bar{\sigma})^3, \bar{\beta}_1 = (\bar{\gamma}_1)^2$ and the empirical value of the kurtosis $\bar{\beta}_2 = \bar{\mu}_4/(\bar{\sigma})^4$ and the excess $\bar{\gamma}_2 = \bar{\beta}_2 - 3$. To simplify the notation the empirical values will be denoted as the theoretical values, respectively. However, one should realize that the estimates which will be determined afterwards may be disturbed by the difference between theoretic and empiric value of the respective moments. $M_{\mathcal{P}}$ estimates with the models $\mathcal{P}_{\text{PD}(l)}$ and $\hat{\mathcal{P}}_{\text{G-C}}$ are compared to the neutral LS-estimates and to the robust Huber M -estimates (see, Eq. (27)). To compare the estimate, we will apply the root-mean-squared (for $\hat{\mathbf{X}} = \hat{\mathbf{X}}_{\text{PD}(l)}, \hat{\mathbf{X}}_{\text{G-C}}, \hat{\mathbf{X}}_H$)

$$\text{RMS}_{\hat{\mathbf{X}}} = [(\hat{\mathbf{X}} - \mathbf{X})^T (\hat{\mathbf{X}} - \mathbf{X}) / r]^{1/2} = (\hat{\mathbf{X}}^T \hat{\mathbf{X}} / 4)^{1/2} \quad (43)$$

To verify $M_{\mathcal{P}}$ estimates from Eq. (16), they are compared with the respective ML-estimates which are computed on the basis of the algorithm in Eqs. (9) and (10).

The very important problem which concerns the Huber method (M_H) is to choose the value of l that defines the interval $\Delta \bar{v} = \langle -l, l \rangle$ for acceptable random errors \bar{v} . It is especially significant in the case of asymmetric distributions. For example, if the distribution has a positive asymmetry coefficient, then some rather large positive errors should be acceptable as random ones (a right heavy tail which is acceptable), thus they should not be treated as gross errors. Figure 9 shows six chosen histograms of the simulated observations (such sets will be applied to show the properties of $M_{\mathcal{P}}$ estimates). Such observations are not affected by gross errors, thus LS-estimates and M_H estimates of the parameter \mathbf{X} should be equal to each other. The analyses, which are obtained for the vectors \mathbf{y}_1 and \mathbf{y}_3 , show that the equality $\hat{\mathbf{X}}_{\text{LS}} = \hat{\mathbf{X}}_{M_H}$ holds for $l \geq 3$ (Table 1). Despite some disturbances of the vector $\hat{\mathbf{X}}_{M_H}$, the value $l = 2.5$, which is often

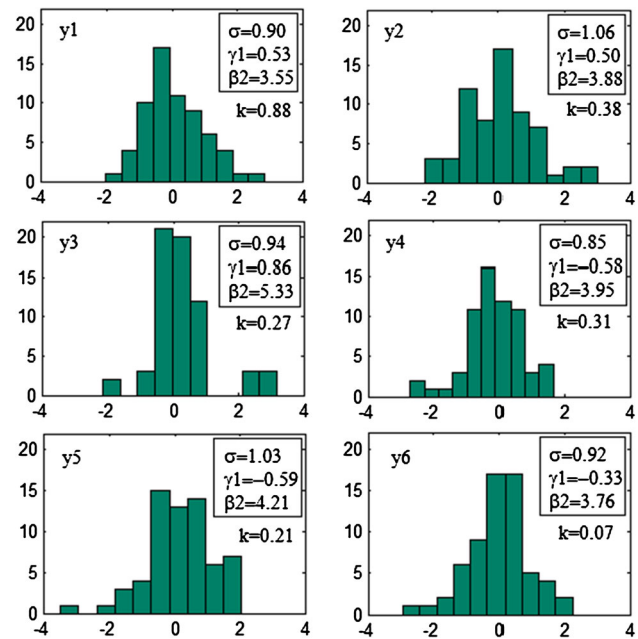


Fig. 9 Histograms of the observations forming the vectors $\mathbf{y}_1, \dots, \mathbf{y}_6$

chosen in practice, can also be acceptable (this value will be chosen to the numerical tests). The smaller values of l might lead to unacceptable M_H estimates of the parameter \mathbf{X} (arbitrary reduction of the influence of some observations which is not justified by the probabilistic model).

The new methods are supposed to be robust against outliers thus they are compared with the method which belongs to the class of robust M -estimation. Tests were carried out for many simulated sets of observations (usually $s = 4, n = 64$), including the sets which were also affected by gross errors. Mostly, it holds that $\text{RMS}_{\hat{\mathbf{X}}_{\text{PD}(l)}} < \text{RMS}_{\hat{\mathbf{X}}_{\text{G-C}}} < \text{RMS}_{\hat{\mathbf{X}}_{\text{LS}}}$, and when gross errors occur $\text{RMS}_{\hat{\mathbf{X}}_{\text{PD}(l)}} < \text{RMS}_{\hat{\mathbf{X}}_H}$ (in each case $M_{\text{PD}(l)}$ and $M_{\text{G-C}}$ estimates with the correction of the weight function for $\epsilon_v = 0.1\sigma$ were equal to the corresponding ML-estimates). Note that if the observation sets are free of gross errors and have small asymmetry then $\hat{\mathbf{X}}_{\text{PD}(l)} = \hat{\mathbf{X}}_{\text{G-C}} = \hat{\mathbf{X}}_{\text{LS}}$ (regardless of the value of the kurtosis), which is in line with theoretical predictions. The detailed results of the tests for the simulated vectors $\mathbf{y}_1, \dots, \mathbf{y}_6$ are presented in Table 2. Gross errors of $g = 10$ and $g = 20$ affected the first measurement of the observation sets.

The $M_{\text{PD}(l)}$ estimates obtained are closer to the theoretical assumed values of the vector $\mathbf{X} = \mathbf{0}$ than LS- and M_H estimates. The approximating model $\hat{\mathcal{P}}_{\text{G-C}} = \{\tilde{f}(v)_4\}$ is not promising here, especially in the case of the observation sets which are affected by gross errors. This is due to the already mentioned properties of the influence function and the weight function of the method (for some values of v , they both tend to the corresponding functions of the neutral LS-estimation). It is also worth noting that $M_{\text{PD}(l)}$ estimates are more robust

Table 1 Disturbances of M_H estimates resulting from the chosen l which defines the interval $\Delta\bar{v} = \langle -l, l \rangle$ for acceptable random errors \bar{v}

	For y_1					For y_3					
	LS	M_H				LS	M_H				
		$l = 1.5$	$l = 2.0$	$l = 2.5$	$l = 3.0$		$l = 1.5$	$l = 2.0$	$l = 2.5$	$l = 3.0$	$l = 3.5$
\hat{H}_A	-0.331	-0.211	-0.252	-0.296	-0.331	-0.271	0.034	-0.080	-0.214	-0.267	-0.331
\hat{H}_B	-0.157	-0.123	-0.130	-0.148	-0.157	-0.128	-0.045	-0.077	-0.112	-0.126	-0.157
\hat{H}_C	-0.083	0.039	-0.026	-0.074	-0.083	-0.062	0.048	0.002	0.040	-0.578	-0.083
\hat{H}_D	-0.160	-0.112	-0.133	-0.151	-0.160	-0.398	0.068	-0.129	-0.307	-0.381	-0.160
$\text{RMS}_{\hat{X}}$	0.204	0.134	0.157	0.186	0.204	0.251	0.050	0.085	0.196	0.243	0.251

Table 2 Comparison $M_{\text{PD}(l)}$ and $M_{\text{G-C}}$ estimates with neutral LS- and robust M_H estimates

	$g = 0$				$g = 10$				$g = 20$			
	LS	M_H	$M_{\text{PD}(l)}$	$M_{\text{G-C}}$	LS	M_H	$M_{\text{PD}(l)}$	$M_{\text{G-C}}$	LS	M_H	$M_{\text{PD}(l)}$	$M_{\text{G-C}}$
y_1												
\hat{H}_A	-0.33	-0.30	-0.16	-0.23	0.22	-0.28	-0.09	0.27	0.76	-0.28	-0.11	0.83
\hat{H}_B	-0.16	-0.15	-0.05	-0.15	-0.04	-0.14	-0.04	-0.05	0.08	-0.14	-0.04	0.08
\hat{H}_C	-0.08	-0.07	0.07	0.01	-0.04	-0.07	0.07	0.04	0.01	-0.07	0.07	0.08
\hat{H}_D	-0.16	-0.15	-0.03	-0.12	-0.04	-0.15	-0.01	-0.03	0.08	-0.15	-0.02	0.06
$\text{RMS}_{\hat{X}}$	0.20	0.19	0.09	0.15	0.11	0.18	0.06	0.14	0.39	0.18	0.07	0.42
y_2												
\hat{H}_A	-0.17	-0.14	-0.02	-0.04	0.37	-0.19	-0.02	0.40	0.92	-0.18	-0.05	0.89
\hat{H}_B	-0.28	-0.28	-0.25	-0.30	-0.16	-0.29	-0.25	-0.19	-0.04	-0.29	-0.25	-0.07
\hat{H}_C	-0.18	-0.20	-0.12	-0.20	-0.13	-0.20	-0.12	-0.15	-0.08	-0.20	-0.12	-0.10
\hat{H}_D	-0.29	-0.26	-0.16	-0.20	-0.17	-0.27	-0.17	-0.09	-0.05	-0.27	-0.17	0.02
$\text{RMS}_{\hat{X}}$	0.24	0.24	0.16	0.20	0.23	0.24	0.16	0.24	0.46	0.24	0.17	0.44
y_3												
\hat{H}_A	-0.27	-0.21	0.06	-0.03	0.28	-0.20	0.08	0.31	0.82	-0.18	0.07	0.70
\hat{H}_B	-0.13	-0.11	-0.01	-0.10	-0.01	-0.11	-0.01	-0.02	0.11	-0.10	-0.01	0.06
\hat{H}_C	-0.06	0.04	0.05	-0.01	-0.01	-0.04	0.06	0.04	0.03	-0.04	0.05	0.04
\hat{H}_D	-0.40	-0.30	0.03	0.02	-0.28	-0.32	0.04	0.13	-0.16	-0.31	0.04	0.07
$\text{RMS}_{\hat{X}}$	0.25	0.20	0.04	0.05	0.20	0.20	0.05	0.17	0.42	0.20	0.05	0.36
y_4												
\hat{H}_A	0.38	0.38	0.28	0.34	0.92	0.41	0.37	0.75	1.47	0.40	0.32	1.22
\hat{H}_B	0.07	0.07	-0.12	-0.06	0.19	0.08	-0.10	0.04	0.31	0.08	-0.11	0.16
\hat{H}_C	0.16	0.16	0.03	0.08	0.21	0.16	0.04	0.12	0.26	0.16	0.03	0.16
\hat{H}_D	0.13	0.13	-0.04	0.02	0.25	0.14	-0.02	0.13	0.37	0.14	-0.03	0.24
$\text{RMS}_{\hat{X}}$	0.22	0.22	0.16	0.18	0.50	0.24	0.19	0.39	0.78	0.23	0.17	0.64
y_5												
\hat{H}_A	0.22	0.22	0.14	0.12	0.76	0.30	0.26	0.61	1.31	0.29	0.21	1.12
\hat{H}_B	0.36	0.36	0.27	0.26	0.48	0.38	0.30	0.37	0.60	0.38	0.29	0.48
\hat{H}_C	-0.05	-0.05	-0.08	-0.012	0.00	-0.04	-0.07	-0.08	0.05	-0.04	-0.07	-0.03
\hat{H}_D	-0.03	-0.03	-0.09	-0.10	0.09	0.01	-0.06	-0.02	0.21	-0.02	-0.07	0.10
$\text{RMS}_{\hat{X}}$	0.21	0.21	0.16	0.16	0.45	0.24	0.20	-0.36	0.73	0.24	0.19	0.61

Table 2 continued

	$g = 0$				$g = 10$				$g = 20$			
	LS	M_H	$M_{PD(l)}$	M_{G-C}	LS	M_H	$M_{PD(l)}$	M_{G-C}	LS	M_H	$M_{PD(l)}$	M_{G-C}
y_6												
\hat{H}_A	0.11	0.11	0.07	0.05	0.66	0.08	0.06	0.55	1.20	0.09	0.03	1.09
\hat{H}_B	-0.06	-0.06	-0.07	-0.10	0.06	-0.07	-0.07	0.01	0.17	-0.07	-0.08	0.10
\hat{H}_C	0.11	0.11	0.08	0.08	0.16	0.11	0.08	0.13	0.21	0.11	0.08	0.17
\hat{H}_D	0.19	0.19	0.09	0.10	0.30	0.18	0.09	0.22	0.42	0.18	0.08	0.35
$RMS_{\hat{X}}$	0.12	0.12	0.08	0.08	0.37	0.12	0.08	0.30	0.65	0.12	0.07	0.58

Table 3 Comparison of ratios $r_H = RMS_{\hat{X}_H}/RMS_{\hat{X}_{LS}}$ and $r_{PD} = RMS_{\hat{X}_{PD(l)}}/RMS_{\hat{X}_{LS}}$

	y_1		y_2		y_3		y_4		y_5		y_6	
	r_H	$r_{PD(l)}$	r_H	$r_{PD(l)}$	r_H	$r_{PD(l)}$	r_H	$r_{PD(l)}$	r_H	$r_{PD(l)}$	r_H	$r_{PD(l)}$
$g = 0$	0.95	0.45	1.00	0.66	0.80	0.16	1.00	0.73	1.00	0.76	1.00	0.67
$g = 10$	1.64	0.54	1.04	0.69	1.00	0.25	0.48	0.38	0.53	0.44	0.32	0.22
$g = 20$	0.46	0.18	0.52	0.37	0.48	0.12	0.29	0.22	0.33	0.26	0.18	0.11

Table 4 Estimates of the coefficient σ_0

	$\hat{\sigma}_{0,\psi}$											
	$g = 0$				$g = 10$				$g = 20$			
	LS	M_H	$M_{PD(l)}$	M_{G-C}	LS	M_H	$M_{PD(l)}$	M_{G-C}	LS	M_H	$M_{PD(l)}$	M_{G-C}
y_1	1.01	0.99	1.07	1.01	1.77	1.04	1.20	1.74	3.03	1.03	1.25	3.00
y_2	1.01	1.01	1.05	1.02	1.71	1.02	1.13	1.69	2.77	1.01	1.15	2.75
y_3	1.04	1.04	0.97	0.95	1.75	1.04	1.04	1.72	2.93	1.02	1.05	2.90
y_4	1.01	1.01	1.02	1.01	1.77	1.06	1.21	1.77	3.10	1.06	1.19	3.10
y_5	1.02	1.02	1.12	1.04	1.51	1.06	1.24	1.53	2.55	1.06	1.23	2.58
y_6	1.02	1.02	1.03	1.02	1.82	1.06	1.14	1.82	3.06	1.05	1.14	3.06

than M_H estimates for $l = 2.5$. This becomes illuminating if one compares the ratios $r_H = RMS_{\hat{X}_H}/RMS_{\hat{X}_{LS}}$ and $r_{PD(l)} = RMS_{\hat{X}_{PD(l)}}/RMS_{\hat{X}_{LS}}$, which are listed in Table 3. In one particular case, in which the gross error was relatively small, M_H estimate was worse than the conventional LS-estimate.

Now let us compare the estimates $\hat{\sigma}_{0,\psi}$ of the standard deviation coefficient obtained for the methods compared here (their values are presented in Table 4). The results obtained are similar to the results of the estimation of the parameter vector of the functional model. Thus, in the case of M_{G-C} method, the estimate $\hat{\sigma}_{0,\psi} = [\hat{\mathbf{v}}^T \mathbf{w}(\mathbf{v}) \hat{\mathbf{v}} / (n - r)]^{1/2}$ is very similar to the non-robust estimate with the weight function $\mathbf{w}(\mathbf{v}) = \sigma^{-2} \mathbf{I}_n$. On the other hand, the estimates obtained in M_H and $M_{PD(l)}$ methods are similar to each other. However, one should expect the larger values of the estimate for larger values of the gross error in the case of $M_{PD(l)}$ method

(in comparison to the theoretical value $\sigma_0 = 1$). To verify such tendency, another test was carried out. It was assumed a smaller observation set ($s = 2, n = 32$) with the moderate asymmetry and excess ($\sigma = 1.19, \gamma_1 = -0.18, \gamma_2 = 0.18$). Thus, we limited redundancy of the observations as well as the distribution anomalies. Such assumptions allow us to avoid blurring of the influence of gross errors on the estimation results. The gross errors of the values $g = 10, g = 20, g = 40$ and $g = 80$ affected the first measurement (as it was previously). The test results show that the estimate of the coefficient σ_0 obtained by applying $M_{PD(l)}$ method is a little bit worse than that obtained in the case of M_H method (see, Fig. 10). However, such difference is not growing with the increase of the value of the gross error.

Such specific properties of $M_{PD(l)}$ estimates (consideration of the known asymmetry and excess) seems substantial, however, in some particular cases it would be more impor-

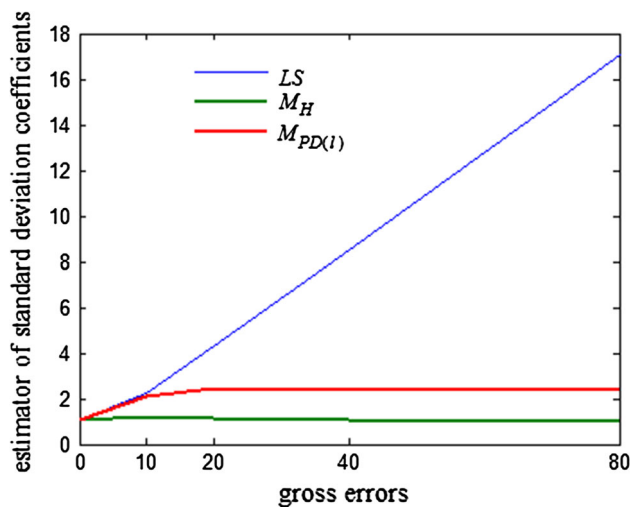


Fig. 10 Influence of gross errors on the estimate of the coefficient σ_0 in LS, M_H and $M_{PD(l)}$ methods (the additional observation set for which $\sigma = 1.19$, $\gamma_1 = -0.18$, $\gamma_2 = 0.18$)

Table 5 $M_{PD(l)}$ estimates $\hat{\sigma}_{0,\psi}$ with natural and with artificial values of the kurtosis

	$\hat{\sigma}_{0,\psi}$				
	$g = 0$	$g = 10$	$g = 20$	$g = 40$	$g = 80$
$\beta_2 = 3.16$ (natural)	1.09	2.13	2.42	2.40	2.36
$\beta_2 = 6.00$ (artificial)	1.09	1.35	1.36	1.36	1.36

tant to make the estimate of \mathbf{X} be highly robust (and hence also the estimate of the variance coefficient from Eq. (18)). $M_{PD(l)}$ estimate can be more robust if we increase artificially the value of the kurtosis. Thus, let us consider the previously used observation set for which $\beta_2 = 3.18$ ($s = 2$, $n = 32$, $\sigma = 1.19$, $\gamma_1 = -0.18$, $\gamma_2 = 0.18$), and the same values of the gross error. Then, one can obtain the results for which $\text{RMS}_{\hat{\mathbf{X}}_{PD(l)}}$ and $\hat{\sigma}_{0,\psi}$ are presented in Fig. 11 (for $3.18 \leq \beta_2 \leq 6.00$) and in Table 5, respectively.

5 Conclusions

The Pearson distributions of types IV and VII are regarded as probabilistic models of $M_{\mathcal{P}}$ estimation which is discussed in the paper. Such variant of $M_{\mathcal{P}}$ estimation is denoted as $M_{PD(l)}$. The platykurtic distributions of types I and II seem also theoretically interesting. However, application of those distributions as probabilistic models of observation errors should be justified convincingly, and the respective optimization procedure of $M_{PD(p)}$ estimation should be proposed. In the case of the probabilistic models with bounded domains, one should apply appropriate objective functions with constraints. In the theoretical part of the

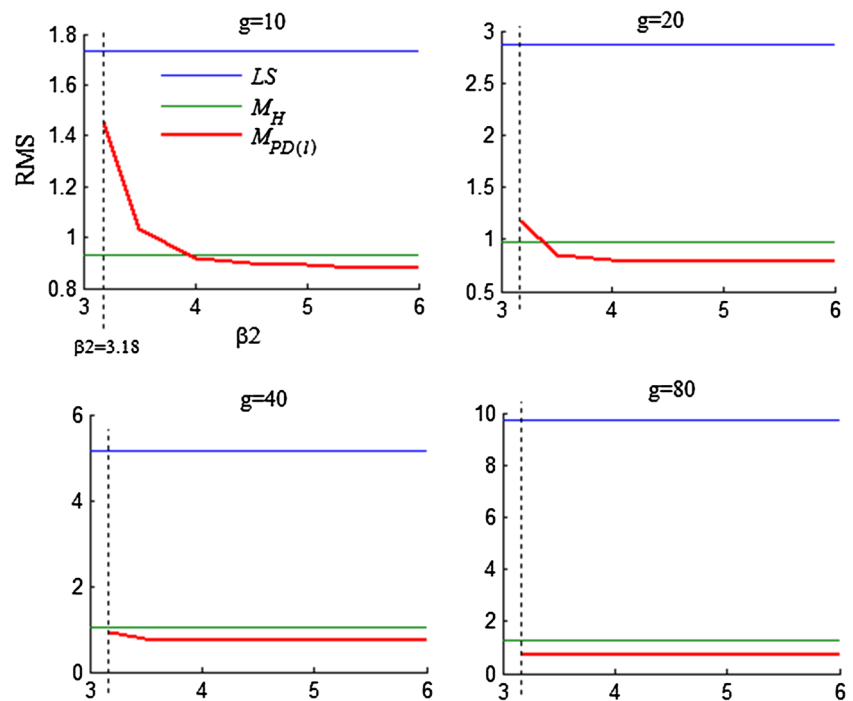
paper we paid attention only to the weight functions of $M_{\mathcal{P}}$ estimations with such models. The distributions PIV and PVII, for which the values of the asymmetry coefficient and the excess are acceptable, were assumed to be an alternative for some distributions which belong to the set W_v . We also considered that a non-Gaussian probability density function can be approximated by series, first of all by the Gram–Charlier series. The influence functions and the weight functions derived from the approximated PDF are the basis for $M_{\mathcal{P}}$ estimation which is denoted as M_{G-C} .

Considering the influence and weight functions, both $M_{PD(l)}$ and M_{G-C} estimation can be regarded as robust M -estimations. Thus, the estimates obtained by applying those methods (the estimates of the parameters of the functional model and the estimates of the variance coefficient) were compared with LS-estimates and the robust estimates of the Huber method (M_H). The theoretical analysis show that M_{G-C} formally belongs to the class of robust estimations, however, for the larger values of gross errors, the robustness of the method is lost. The conclusion is also confirmed by the numerical tests. It follows that M_{G-C} estimation is only “a bit” robust against gross errors. For example, if the gross error is equal to $g = 20$, then the ratio $\text{RMS}_{\hat{\mathbf{X}}_{G-C}}/\text{RMS}_{\hat{\mathbf{X}}_{LS}}$ is within the range $0.82 \div 1.07$. Similar ratio computed for $M_{PD(l)}$ estimates is within the range $0.11 \div 0.37$. The results of M_{G-C} estimation are also not promising for the sets which are not affected by gross errors (in comparison with $M_{PD(l)}$ estimates). In such case, the ratio $\text{RMS}_{\hat{\mathbf{X}}_{G-C}}/\text{RMS}_{\hat{\mathbf{X}}_{LS}}$ is within the range $0.20 \div 0.83$ (and $0.16 \div 0.76$ for $M_{PD(l)}$ estimates). If the further terms of the G–C series were applied, one would obtain better results. However, in such a case the higher order moments, namely μ_5 , μ_6 etc. should be known. Such moments are rarely determined for sets of geodetic observations so far.

$M_{PD(l)}$ estimates are closer to the true values of the parameter $\mathbf{X} = \mathbf{0}$ than the corresponding LS-estimates in most cases which are presented in the paper (Sect. 4 shows several such examples, see Tables 2 and 3). This conclusion concerns sets affected by gross errors as well as the sets which are free of such errors. One should pay attention to robustness of $M_{PD(l)}$ estimates. Comparing those estimates with the Huber estimates (for rather rigorous value of $l = 2.5$), we can say that the ratio $\text{RMS}_{\hat{\mathbf{X}}_{PD(l)}}/\text{RMS}_{\hat{\mathbf{X}}_{LS}}$ is smaller than $\text{RMS}_{\hat{\mathbf{X}}_H}/\text{RMS}_{\hat{\mathbf{X}}_{LS}}$. The tests presented show that for the gross error of $g = 20$ we obtained the ratio values within the range $0.18 \div 0.52$ for M_H estimates and $0.11 \div 0.37$ for $M_{PD(l)}$ estimates.

Estimation of the variance coefficients plays a major role in quality assessment of the adjustment results (Sect. 4 presents the values of the respective estimates). Robustness of the estimate $\hat{\sigma}_{0,\psi}^2$ seems to be important from a theoretical as well as a practical point of view. For the sets which

Fig. 11 Growing robustness of $M_{PD(l)}$ estimate of \mathbf{X} by artificial increasing of the kurtosis



are not affected by gross errors, the values of the estimates of the variance coefficient obtained for the methods compared are similar (see, Table 4 for $g = 0$). If gross errors occur, then the estimates of the coefficient σ_0^2 obtained in the case of $M_{PD(l)}$ method are a little bit larger than those for M_H estimation (however, they are always much smaller than non-robust estimates obtained with application of the weight function $\mathbf{w}(\mathbf{v}) = \sigma^{-2}\mathbf{I}_n$). The special tests show that the estimates $\hat{\sigma}_{0,\psi}^2$ obtained in $M_{PD(l)}$ method do not get worse with the growing value of the gross error.

The robustness of $M_{PD(l)}$ estimates is related to value of the kurtosis. Remember that M_P method is a weak estimation in the case of $\beta_2 < 3$ (platykurtic distributions) and the probabilistic model $\mathcal{P}_{PD(p)} = \{\text{PI}, \text{PII}\}$. For $\beta_2 = 3$ and the model $\mathcal{P}_{PD(m)} \equiv \mathcal{P}_{ND}$ (LS-method), it becomes a neutral estimation. Robustness of the method increases with the growing value of the kurtosis (if the asymmetry coefficient is the same). Thus, if $\beta_1 = 0$, then $M_{PD(l)}$ estimation can be regarded as a robust M -estimation which robustness is adjusted by β_2 . In such a case the kurtosis becomes only a parameter which controls the robustness of the method (pseudo-kurtosis); hence it does not have to be related to the real values of the second and fourth moments. In such an approach, the Gaussian distributions with an unacceptable admixture are replaced with the Pearson distributions with the equivalent kurtosis which is artificially overestimated. Such replacement of the distribution is just formal, however, it might be of importance to statistical interpretation of the results (for example, it can be done by applying appropriate statistical tests, which requires some additional theoretical

researches and analyses). In respect of robust M -estimation, such an approach is related to changes in the influence function and the weight function which are easy to be planned.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Andreassen TW, Lorentzen BG, Olsson H (2006) The impact of non-normality and estimation methods in SEM on satisfaction research in marketing. *Qual Quant* 40:39–58
- Baarda W (1968) A test procedure for use in geodetic networks. *Neth Geod Comm Publ Geod, New Ser* 2(5):27–55
- Berberan-Santos MN (2007) Expressing a probability density function in terms of another PDF: a generalized Gram–Charlier expansion. *J Math Chem* 42:585–594
- Cen M, Li Z, Ding X, Zhuo J (2003) Gross error diagnostics before least squares adjustment of observations. *J Geod* 77:503–513
- Charlier CVL (1906) Researches into the theory of probability. Håkan Ohlsson, Lund
- Crawford JR, Walsh JE (1960) Empirical examination of Edgeworth series. *Ann Inst Stat Math* 1:13–26
- Daniels HE (1954) Saddlepoint approximations in statistics. *Ann Math Stat* 25(4):631–650
- Duchnowski R (2011) Robustness of the strategy for testing leveling mark stability based on rank tests. *Surv Rev* 43:687–699
- Duchnowski R (2013) Hodges–Lehmann estimates in deformation analysis. *J Geod* 87:873–884
- Duchnowski R, Wiśniewski Z (2011) Estimation of the shift between parameters of functional models of geodetic observations by applying M_{split} estimation. *J Surv Eng* 138:1–8

- Dumalski A, Wiśniewski Z (1994) Concept of adjustment method with application of Edgeworth series. *Polish Acad Sci Geod Cartogr* 43:219–237
- Dzhun' IV (1969) Pearson VII type distribution in errors of latitude oscillation observations. *Astrom Astrofiz* 2:101–115
- Dzhun' IV (1991) Pearson VII type distribution of errors of HIS laser observations. *Kinem Fiz Nebes Tel* 7:82–91
- Dzhun' IV (1992) Pearson distribution of type VII used to approximate observation errors in astronomy. *Meas Tech* 35:277–282
- Dzhun' IV (2011) Method for diagnostics of mathematical models in theoretical astronomy and astrometry. *Kinemat Phys Celest Bodies Math Proc Astron Data* 27:260–264
- Dzhun' IV (2012) What should be the observation–calculation residuals in modern astrometric experiments. *Kinemat Phys Celest Bodies Math Proc Astron Data* 28:72–80
- Elderton WP (1953) Frequency curves and correlation. Cambridge University Press, London
- Ferrari D, Yang Y (2010) Maximum L_q -likelihood estimation. *Ann Stat* 38:753–783
- Fiori AM, Zenga M (2009) Karl Pearson and the origin of kurtosis. *Int Stat Rev* 77:40–50
- Gleinsvik P (1971) Zur Leistungsfähigkeit der Methode der kleinsten Quadrate bei Ausgleichung nicht normalverteilter Beobachtungen. Theoretische Untersuchungen. *ZfV*, Heft 6:224–233
- Gleinsvik P (1972) Zur Leistungsfähigkeit der Methode der kleinsten Quadrate bei Ausgleichung nicht normalverteilter Beobachtungen. Experimentelle Prüfung. *ZfV*, Heft 9:385–393
- Gui Q, Zhang J (1998) Robust biased estimation and its applications in geodetic adjustment. *J Geod* 72:430–435
- Hald A (2007) A history of parametric statistical inference from Bernoulli to Fisher, 1713–1935. Sources and studies in the history of mathematics and physical sciences. Springer Science+Business Media, LLC, pp 111–129
- Hampel FR (1974) The influence curve and its role in robust estimation. *J Am Stat Assoc* 69:383–397
- Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (1986) Robust statistics. The approach based on influence functions. Wiley, New York
- Hodges JL, Lehmann EL (1963) Estimates of location based on rank tests. *Ann Math Stat* 34:598–611
- Hu X, Huang C, Liao X (2001) A new solution assessment approach and its application to space geodesy data analysis. *Celest Mech Dyn Astron* 81:265–278
- Huang Y, Mertikas SP (1995) On the design of robust regression estimators. *Manuscr Geod* 20:145–160
- Huber PJ (1981) Robust statistics. The approach based on influence functions. Wiley, New York
- Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106(4):620–630
- Johnson NL (1949) Systems of frequency curves generated by methods of translation. *Biometrika* 36:149–176
- Jones GA, Jones JM (2000) Information and coding theory. Springer undergraduate mathematics series. Springer, Heidelberg
- Kadaj R (1988) Eine verallgemeinerte Klasse von Schätzverfahren mit praktischen Anwendungen. *ZfV* 113:157–166
- Kamiński W, Wiśniewski Z (1994) The method of growing rigor for the adjustment of geodetic observations contaminated by gross errors. *Manuscr Geod* 19:55–61
- Kasiotczuk B (1997) Estimation of asymmetry and kurtosis coefficients in the process of geodetic network adjustment by the least-squares method. *J Geod* 7:131–136
- Klein A, Moosbrugger H (2000) Maximum likelihood of latent interaction effects with the LMS method. *Psychometrika* 65:457–474
- Koch KR (1990) Bayesian inference with geodetic applications. Springer, Berlin
- Koch KR (1996) Robuste Parameterschätzung. *Allg Vermess Nach* 103:1–18
- Kolassa JE (2006) Series approximation methods in statistics. Lecture Notes on Statistics, vol 88. Springer, Berlin, pp 31–62
- Krarup T, Kubik K (1983) The Danish method; experience and philosophy. Deutsche Geodätische Kommission, München, Reihe A, Heft 7:131–134
- Kukuča J (1967) Some problems in estimating the accuracy of a measuring method. *Studia Geoph et Geod* 11:21–33
- Luo X, Mayre M, Heck B (2011) On the probability distribution of GNSS carrier phase observations. *GPS Solut* 15:369–379
- Marshall J (2002) L_1 -norm pre analysis measures for geodetic networks. *J Geod* 76:334–344
- Mooijaart A (1985) Factor analysis for non-normal variables. *Psychometrika* 50:323–342
- Mukhopadhyay N (2005) Dependence or independence of the sample mean and variance in non -IID or non-normal cases and the role of some tests of independence. Recent advances in applied probability. Springer Science + Business Media, Inc., Book Chapter, pp 397–426
- Nagahara Y (2007) A method of calculating the downside risk by multivariate nonnormal distributions. *Asia Pac Finan Mark* 15:175–184
- Nagahara Y (2011) Using nonnormal distributions to analyze the relationship between stock returns in Japan and US. *Asia Pac Finan Mark* 18:429–443
- Pearson K (1920) The fundamental problem of practical statistics. *Statistics* 13:1–16
- Prószynski W (1997) Measuring the robustness potential of the least-squares estimation: geodetic illustration. *J Geod* 71:652–659
- Romanowski M, Green E (1983) Reflexions on the kurtosis of samples of errors. *Bull Géod* 57:62–82
- Serfling R (1980) Approximation theorems of mathematical statistics. Wiley, New York
- Teunissen PJG (1990) Nonlinear least squares. *Manuscr Geod* 15:137–150
- Tomozawa Y (1974) Quasinormal distributions and expansion at the mode. *J Stat Phys* 11:195–204
- Wassef AM (1959) Note of the application of mathematical statistics to the analysis of levelling errors. *Bull Géod* 52:19–26
- Winkelmann R, Boes S (2006) Analysis of microdata. Springer, Berlin
- Winter R (1978) Theoretische und praktische untersuchungen zu modulierten Normalverteilungen nach Romanowski. Wissenschaftliche Arbeiten der Lehrstühle für Geodäsie, Photogrammetrie und Kartographie an der Technischen Universität Hannover
- Wiśniewski Z (1985) The effect of the asymmetry of geodetic observation error distribution on the results of adjustment by least squares method. *Polish Acad Sci Geod Cartogr* 34:11–21
- Wiśniewski Z (1987) Method of geodetic network adjustment in extend to probabilistic measurement error properties. *Sci Bull Staszic Acad Min Metall Geod* 95:73–88
- Wiśniewski Z (1989) The RP method. Part I-III (in Polish). *Geodezja i Kartografia* 38:25–69
- Wiśniewski Z (1991) Comparative categories in analysis of the methods of geodetic observations adjustment. *Sci Bull Staszic Acad Min Metall Geod* 112:41–55
- Wiśniewski Z (1995) Moment vectors and their estimation after least squares adjustment. *Bollettino di Geodesia e Scienze Affini* 54:373–386
- Wiśniewski Z (1996) Estimation of the third and fourth order central moments of measurement errors from sums of powers of least squares adjustment residuals. *J Geod* 70:256–262
- Wiśniewski Z (2009) Estimation of parameters in a split functional model of geodetic observations (M_{split} estimation). *J Geod* 83:105–120
- Wiśniewski Z (2010) $M_{split(q)}$ estimation: estimation of parameters in a multi split functional model of geodetic observations. *J Geod* 84:355–372

- Xi Z, Hu C, Youn BD (2012) A comparative study of probability estimation methods for reliability analysis. *Struct Multidiscip Optim* 45:33–52
- Yang Y (1994) Robust estimation for dependent observations. *Manuscr Geod* 19:10–17
- Yang Y (1997) Estimators of covariance matrix at robust estimation based on influence functions. *ZfV* 122:166–174
- Yang Y (1999) Robust estimation of geodetic datum transformation. *J Geod* 73:268–274
- Yang Y, Cheng MK, Shum CK, Tapley BD (1999) Robust estimation of systematic errors of satellite laser range. *J Geod* 73:345–349
- Yang Y, Song L, Xu T (2002) Robust estimation for correlated observations based on bifactor equivalent weights. *J Geod* 76:353–358
- Youcai H, Mertikas SP (1995) On the design of robust regression estimators. *Manuscr Geod* 20:145–160